Functional central limit theorems for rough volatility

Blanka Horvath[1] · Antoine Jacquier[2] · Aitor Muguruza[3] · Andreas Søjmark[4]

Abstract

The non-Markovian nature of rough volatility makes Monte Carlo methods challenging, and it is in fact a major challenge to develop fast and accurate simulation algorithms. We provide an efficient one for stochastic Volterra processes, based on an extension of Donsker's approximation of Brownian motion to the fractional Brownian case with arbitrary Hurst exponent $H \in (0, 1)$. Some of the most relevant consequences of this 'rough Donsker (rDonsker) theorem' are functional weak convergence results in Skorokhod space for discrete approximations of a large class of rough stochastic volatility models. This justifies the validity of simple and easy-to-implement Monte Carlo methods, for which we provide detailed numerical recipes. We test these against the current benchmark hybrid scheme and find remarkable agreement (for a large range of values of H). Our rDonsker theorem further provides a weak convergence proof for the hybrid scheme itself and allows constructing binomial trees for rough volatility models, the first available scheme (in the rough volatility context) for early exercise options such as American or Bermudan options.

✉ A. Jacquier
a.jacquier@imperial.ac.uk

B. Horvath
blanka.horvath@maths.ox.ac.uk

A. Muguruza
aitor.muguruza-gonzalez15@imperial.ac.uk

A. Søjmark
a.sojmark@lse.ac.uk

[1] Department of Mathematics, University of Oxford, Oxford, UK

[2] Department of Mathematics, Imperial College London, and the Alan Turing Institute, London, UK

[3] Kaiju Capital Management and Department of Mathematics, Imperial College London, London, UK

[4] Department of Statistics, London School of Economics, London, UK

1 Introduction

Fractional Brownian motion has a long and famous history in probability, stochastic analysis and their applications to diverse fields; see Hurst [39], Kolmogorov [44] and Mandelbrot and Van Ness [52]. Recently, it has experienced a new renaissance in the form of fractional volatility models in mathematical finance. These were first introduced by Comte and Renault [18] and later studied theoretically by Djehiche and Eddahbi [20], Alòs et al. [2] and Fukasawa [29], and given financial motivation and data consistency by Gatheral et al. [32] and Bayer et al. [6]. Since then, a vast literature has pushed the analysis in many directions (see Bayer et al. [5], Bayer et al. [7], Bennedsen et al. [11], El Euch and Rosenbaum [24], Forde and Zhang [26], Fukasawa et al. [30], Guennoun et al. [33], Gulisashvili [34], Horvath et al. [38], Jacquier et al. [41] and Neuman and Rosenbaum [58]), leading to theoretical and practical challenges to understand and implement these models. One of the main issues, at least from a practical point of view, is on the numerical side: the absence of Markovianity rules out PDE-based schemes, and simulation is the only possibility. However, classical simulation methods for fractional Brownian motion (based on Cholesky decomposition or circulant matrices) are notoriously slow, and faster techniques are needed. The state-of-the-art so far is the recent hybrid scheme developed by Bennedsen et al. [10] and its turbocharged version in McCrickerd and Pakkanen [54]. We rise here to this challenge and propose an alternative tree-based approach, mathematically rooted in an extension of Donsker's theorem to rough volatility.

Donsker [21] (and later Lamperti [48]) proved a functional central limit for Brownian motion, thereby providing a theoretical justification of its random walk approximation. Many extensions have been studied in the literature, and we refer the interested reader to Dudley [23, Chap. 10] for an overview. In the fractional case, Sottinen [66] and Nieminen [59] constructed – following Donsker's ideas of using i.i.d. sequences of random variables – an approximating sequence converging to fractional Brownian motion with Hurst parameter $H > 1/2$. In order to deal with the non-Markovian behaviour of fractional Brownian motion, Taqqu [67] considered sequences of non-i.i.d. random variables, again with the restriction $H > 1/2$. Unfortunately, these methodologies do not seem to carry over to the 'rough' case $H < 1/2$, mainly because of the topologies involved. The recent development of rough paths theory (see Friz and Victoir [28, Sect. 13.4] and Lyons [51]) provided an appropriate framework to extend Donsker's results to processes with sample paths of Hölder regularity strictly smaller than $1/2$. For $H \in (1/3, 1/2)$, Bardina et al. [3] used rough paths to show that functional central limit theorems (in the spirit of Donsker) apply. This in particular suggests that the natural topology to use for fractional Brownian

motion with $H < 1/2$ is the topology induced by the Hölder norm of the sample paths. Indeed, switching the topology from the Skorokhod one used by Donsker to the (stronger) Hölder topology is the right setting for rough central limit theorems, as we outline in this paper. Recent results by Bender and Parczewski [9] and Parczewski [60, 61] provide convergence for (geometric) fractional Brownian motions with general $H \in (0, 1)$ using Wick calculus, assuming that the approximating sequences are Bernoulli random variables. We extend this (Theorem 2.11) to a universal functional central limit theorem, involving general (discrete or continuous) random variables as approximating sequences, only requiring finiteness of moments.

We consider a general class of continuous processes with any Hölder regularity, including fractional Brownian motion with $H \in (0, 1)$, truncated Brownian semi-stationary processes, Gaussian Volterra processes, as well as rough volatility models recently proposed in the financial literature. The fundamental novelty here is an approximating sequence capable of simultaneously keeping track of the approximated rough volatility process (fractional Brownian motion, Brownian semistationary process, or any continuous path functional thereof) and of the underlying Brownian motion. This is crucial in order to take into account the correlation of the two processes, the so-called leverage effect in financial modelling. While approximations of two-dimensional (correlated) semimartingales are well understood in the standard case, the rough case is so far an open problem. Our analysis easily generalises beyond Brownian drivers to more general semimartingales, emphasising that the subtle, yet essential, difficulties lie in the passage from the semimartingale setup to the rough case. This is the first Monte Carlo method available in the literature, specifically tailored to two-dimensional rough systems, based on an approximating sequence for which we prove a Donsker–Lamperti-type functional central limit theorem (FCLT). This further provides a pathwise justification of the hybrid scheme by Bennedsen et al. [10], and allows the use of tree-based schemes, opening the doors to pricing early-exercise options such as American options. In Sect. 2, we present the class of models we are considering and state our main results. The proof of the main theorem is developed in Sect. 3 in several steps. We reserve Sect. 4 to applications of the main result, namely weak convergence of the hybrid scheme, binomial trees as well as numerical examples. We present simple numerical recipes, providing a pedestrian alternative to the advanced hybrid schemes in [10, 54], and develop a simple Monte Carlo scheme with low implementation complexity, for which we provide comparison charts against [10] in terms of accuracy and against [54] in terms of speed. Reminders on Riemann–Liouville operators and additional technical proofs are postponed to the Appendix.

Notations: For the interval $\mathbb{I} := [0, 1]$, we denote by $\mathcal{C}(\mathbb{I})$ and $\mathcal{C}^\alpha(\mathbb{I})$ the spaces of continuous and α-Hölder-continuous functions on $\mathbb{I}$ with Hölder regularity $\alpha \in (0, 1)$, respectively. We also need $\mathcal{C}^1(\mathbb{I}) := \{f : \mathbb{I} \to \mathbb{R} : f' \text{ exists and is continuous on } \mathbb{I}\}$. For $0 \leq \alpha \leq 1$, we define $\mathcal{C}_+^\alpha(\mathbb{I}) := \{f \in \mathcal{C}^\alpha(\mathbb{I}) : f \geq 0\}$. Both definitions imply bounded first order derivatives on $\mathbb{I}$. We use $C, \widetilde{C}, \widehat{C}, C_1, C_2, \overline{C}, \underline{C}$ as strictly positive real constants which may change from line to line and whose exact values do not matter.

2 Weak convergence of rough volatility models

Donsker's invariance principle, also termed 'functional central limit theorem', ensures the weak convergence in the Skorokhod space of an approximating sequence to a Brownian motion. As opposed to the central limit theorem, Donsker's theorem is a pathwise statement which ensures that convergence takes place for all times. This result is particularly important for Monte Carlo methods which aim to approximate pathwise functionals of a given process (an essential requirement to price path-dependent financial securities for example). We prove here a version of Donsker's result not only in the Skorokhod topology, but also in the stronger Hölder topology, for a general class of continuous stochastic processes.

2.1 Hölder spaces and fractional operators

For $\beta \in (0, 1)$, the β-Hölder space $\mathcal{C}^\beta(\mathbb{I})$ with the norm

$$\|f\|_\beta := |f|_\beta + \|f\|_\infty = \sup_{\substack{t,s \in \mathbb{I} \\ t \neq s}} \frac{|f(t) - f(s)|}{|t - s|^\beta} + \max_{t \in \mathbb{I}} |f(t)|,$$

is a non-separable Banach space; see Krylov [45, Chap. 3]. In the spirit of Riemann–Liouville fractional operators recalled in Appendix A, we introduce generalised fractional operators. For $\lambda \in (0, 1)$ and any $\alpha \in (-\lambda, 1 - \lambda)$, we define the space $\mathcal{L}^\alpha := \{g \in \mathcal{C}^2((0,1]) : |\frac{g(u)}{u^\alpha}|, |\frac{g'(u)}{u^{\alpha-1}}|$ and $|\frac{g''(u)}{u^{\alpha-2}}|$ are bounded$\}$. When $\alpha > 0$, we define the function g at the origin to be $g(0) = 0$ by continuous extension.

Definition 2.1 For any $\lambda \in (0, 1)$ and $\alpha \in (-\lambda, 1 - \lambda)$, the *generalised fractional operator (GFO)* associated to $g \in \mathcal{L}^\alpha$ is defined on $\mathcal{C}^\lambda(\mathbb{I})$ as

$$(\mathcal{G}^\alpha f)(t) := \begin{cases} \int_0^t (f(s) - f(0)) \frac{\mathrm{d}}{\mathrm{d}t} g(t - s) \mathrm{d}s, & \text{if } \alpha \in (0, 1 - \lambda), \\ \frac{\mathrm{d}}{\mathrm{d}t} (\int_0^t (f(s) - f(0)) g(t - s) \mathrm{d}s), & \text{if } \alpha \in (-\lambda, 0). \end{cases} \tag{2.1}$$

We further use the notation $G(t) := \int_0^t g(u) \mathrm{d}u$ for any $t \in \mathbb{I}$. Of particular interest in mathematical finance are the kernels

Riemann–Liouville:	$g(u) = u^\alpha$	for $\alpha \in (-1, 1)$;
Gamma fractional:	$g(u) = u^\alpha e^{\beta u}$	for $\alpha \in (-1, 1), \beta < 0$;
power-law:	$g(u) = u^\alpha (1 + u)^{\beta - \alpha}$	for $\alpha \in (-1, 1), \beta < -1$.

$$\tag{2.2}$$

The result below generalises the classical mapping properties of Riemann–Liouville fractional operators first proved by Hardy and Littlewood [36] and is of fundamental importance in the rest of our analysis.

Proposition 2.2 *For any $\lambda \in (0, 1)$ and $\alpha \in (-\lambda, 1 - \lambda)$, the operator $\mathcal{G}^\alpha$ is continuous from $\mathcal{C}^\lambda(\mathbb{I})$ to $\mathcal{C}^{\lambda+\alpha}(\mathbb{I})$.*

The proof can be found in Appendix C. This result is analogous to the classical Schauder estimates, phrased in terms of convolution with a suitable regularising kernel, as e.g. treated in Friz and Hairer [27, Theorem 14.17] and Broux et al. [15, Theorem 2.13 and Lemma 2.9], but in settings that are slightly different from ours.

We develop here an approximation scheme for the following system which generalises the concept of rough volatility introduced in Alòs et al. [2], Fukasawa [29] and Gatheral et al. [32] in the context of mathematical finance. The process X represents the dynamics of the logarithm of a stock price process, under a given reference pricing measure, with

$$
\begin{cases}
\mathrm{d}X_t = -\frac{1}{2}V_t\mathrm{d}t + \sqrt{V_t}\,\mathrm{d}B_t, & X_0 = 0, \\
V_t = \Phi(\mathcal{G}^\alpha Y)(t),
\end{cases}
\tag{2.3}
$$

with $\alpha \in (-\frac{1}{2}, \frac{1}{2})$ and where Y is the (strong) solution to the stochastic differential equation

$$
\mathrm{d}Y_t = b(Y_t)\mathrm{d}t + a(Y_t)\mathrm{d}W_t, \qquad Y_0 \in \mathcal{D}_Y,
\tag{2.4}
$$

where $\mathcal{D}_Y$ denotes the state space of Y, usually $\mathbb{R}$ or $\mathbb{R}_+$. The two Brownian motions B and W are defined on a common filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{I}}, \mathbb{P})$ and are correlated by the parameter $\rho \in [-1, 1]$, and we introduce $\overline{\rho} := \sqrt{1 - \rho^2}$. We further let Φ be any operator such that for all $\gamma \in (0, 1)$, we have $\Phi : \mathcal{C}^\gamma(\mathbb{I}) \to \mathcal{C}_+^\gamma(\mathbb{I})$ with Φ continuous from $(\mathcal{C}^\gamma(\mathbb{I}), \|\cdot\|_\gamma)$ to itself. This in particular ensures that whenever $Y \in \mathcal{C}^\lambda(\mathbb{I})$, then $V \in \mathcal{C}_+^{\alpha+\lambda}(\mathbb{I})$, i.e., V is nonnegative and belongs to $\mathcal{C}^{\alpha+\lambda}(\mathbb{I})$. As an example, one can consider a so-called Nemyckij operator $\Phi(f) := \phi \circ f$ given by composition with some $\phi : \mathbb{R} \to \mathbb{R}_+$, in which case Drábek [22] has shown that the operator Φ is continuous from $(\mathcal{C}^\gamma(\mathbb{I}), \|\cdot\|_\gamma)$ to $(\mathcal{C}^\gamma(\mathbb{I}), \|\cdot\|_\gamma)$ for all $\gamma \in (0, 1)$ if and only if $\phi \in \mathcal{C}^1(\mathbb{R})$. It remains to formulate a precise definition for $\mathcal{G}^\alpha W$ (Proposition 2.4) and for $\mathcal{G}^\alpha Y$ (Corollary 2.5) to fully specify the system (2.3) and clarify the existence of solutions.

Assumption 2.3 There exist $C_b, C_a > 0$ such that for all $y \in \mathcal{D}_Y$,

$$
|b(y)| \leq C_b(1 + |y|), \qquad |a(y)| \leq C_a(1 + |y|),
$$

where a and b are continuous functions such that there is a unique strong solution to (2.4).

Existence of solutions to (2.4) along with Assumption 2.3 need to be checked case by case beyond standard Lipschitz and linear growth conditions (we provide a showcase of models in Examples 2.6–2.9 below satisfying existence and pathwise uniqueness conditions). Not only is the solution to (2.4) continuous, but $(\frac{1}{2} - \varepsilon)$-Hölder-continuous for any $\varepsilon \in (0, \frac{1}{2})$ as a consequence of the Kolmogorov–Čentsov theorem in Čentsov [16]. Existence and precise meaning of $\mathcal{G}^\alpha Y$ are delicate and treated below.

2.2 Examples

Before constructing our approximation scheme, let us discuss a few examples of processes within our framework. As a first useful application, these generalised fractional operators provide a (continuous) mapping between a standard Brownian motion and its fractional counterpart.

Proposition 2.4 *For any* $\alpha \in (-\frac{1}{2}, \frac{1}{2})$, *the equality* $(\mathcal{G}^\alpha W)(t) = \int_0^t g(t-s)\mathrm{d}W_s$ *holds almost surely for all* $t \in \mathbb{I}$.

Proof Since the paths of Brownian motion are $(\frac{1}{2} - \varepsilon)$-Hölder-continuous for any $\varepsilon \in (0, \frac{1}{2})$, existence (and continuity) of $\mathcal{G}^\alpha W$ is guaranteed for all $\alpha \in (-\frac{1}{2}, \frac{1}{2})$. When $\alpha \in (0, \frac{1}{2})$, the kernel is smooth and square-integrable, so that Itô's product rule yields by Proposition 2.2 (since $g(0) = 0$ and $\alpha > 0$) that

$$(\mathcal{G}^\alpha W)(t) = \int_0^t (W_s - W_0) \frac{\mathrm{d}}{\mathrm{d}t} g(t-s)\mathrm{d}s$$

$$= g(t)(W_0 - W_0) - g(0)(W_t - W_0) + \int_0^t g(t-s)\mathrm{d}W_s$$

$$= \int_0^t g(t-s)\mathrm{d}W_s,$$

and the claim holds. For $\alpha \in (-\frac{1}{2}, 0)$ and any $\varepsilon > 0$, introduce the operator

$$(\mathcal{G}_\varepsilon^{1+\alpha} f)(t) := \int_0^{t-\varepsilon} g(t-s)\big(f(s) - f(0)\big)\mathrm{d}s \qquad \text{for all } t \in \mathbb{I},$$

which satisfies $\frac{\mathrm{d}}{\mathrm{d}t} \lim_{\varepsilon\downarrow 0}(\mathcal{G}_\varepsilon^{1+\alpha} f)(t) = (\mathcal{G}^\alpha f)(t)$ pointwise. Now for any $t \in \mathbb{I}$, almost surely,

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathcal{G}_\varepsilon^{1+\alpha} W)(t) = g(\varepsilon)(W_{t-\varepsilon} - W_0) - g(t)(W_0 - W_0) + \int_0^{t-\varepsilon} W_s \frac{\mathrm{d}}{\mathrm{d}t} g(t-s)\mathrm{d}s$$

$$= g(\varepsilon)W(0) + \int_0^{t-\varepsilon} g(t-s)\mathrm{d}W_s. \tag{2.5}$$

Then as ε tends to zero, the right-hand side of (2.5) tends to $\int_0^t g(t-s)\mathrm{d}W_s$ and furthermore, the convergence is uniform. On the other hand, the equalities

$$(\mathcal{G}_0^{1+\alpha} W)(t) - (\mathcal{G}_0^{1+\alpha} W)(0) = \lim_{\varepsilon\downarrow 0} \big((\mathcal{G}_\varepsilon^{1+\alpha} W)(t) - (\mathcal{G}_\varepsilon^{1+\alpha} W)(0)\big)$$

$$= \lim_{\varepsilon\downarrow 0} \int_0^t \left(\frac{\mathrm{d}}{\mathrm{d}s} \mathcal{G}_\varepsilon^{1+\alpha} W\right)(s)\mathrm{d}s$$

$$= \int_0^t \lim_{\varepsilon\downarrow 0} \left(\frac{\mathrm{d}}{\mathrm{d}s} \mathcal{G}_\varepsilon^{1+\alpha} W\right)(s)\mathrm{d}s$$

$$= \int_0^t \left(\int_0^s g(s-u)\mathrm{d}W_u\right)\mathrm{d}s$$

hold since convergence is uniform on compacts, and the fundamental theorem of calculus concludes the proof. □

Minimal changes to the proof of Proposition 2.4 also yield the following result.

Corollary 2.5 *If Y solves* (2.4)*, then* $(\mathcal{G}^\alpha Y)(t) = \int_0^t g(t-s) \mathrm{d}Y_s$ *almost surely for all* $t \in \mathbb{I}$ *and* $\alpha \in (-\frac{1}{2}, \frac{1}{2})$*.*

Up to a constant multiplicative factor C_α, the (left) fractional Riemann–Liouville operator (see Appendix A) is identical to the GFO in (2.2) so that the Riemann–Liouville (or type-II) fractional Brownian motion can be written as $C_\alpha \mathcal{G}^\alpha W$. Proposition 2.2 then implies that the Riemann–Liouville operator is continuous from $\mathcal{C}^{1/2}(\mathbb{I})$ to $\mathcal{C}^{1/2+\alpha}(\mathbb{I})$ for $\alpha \in (-\frac{1}{2}, \frac{1}{2})$. Each kernel in (2.2) gives rise to processes proposed by Barndorff-Nielsen and Schmiegel [4] for turbulence and financial modelling.

Example 2.6 The rough Bergomi model by Bayer et al. [6] reads

$$V_t = \xi_0(t)\mathcal{E}\left(2\nu C_H \int_0^t (t-s)^\alpha \mathrm{d}W_s\right),$$

with $V_0, \nu, \xi_0(\cdot) > 0$, $\alpha \in (-\frac{1}{2}, \frac{1}{2})$ and $\mathcal{E}(\cdot)$ is the stochastic exponential. (We remark that this coincides here with the Wick stochastic exponential.) This corresponds exactly to (2.3) with $g(u) = u^\alpha$, $Y = W$ and

$$\Phi(\varphi)(t) := \xi_0(t)\exp\left(2\nu C_H \varphi(t)\right)\exp\left(-2\nu^2 C_H^2 \int_0^t (t-s)^{2\alpha} \mathrm{d}s\right).$$

Example 2.7 A truncated Brownian semistationary (TBSS) process is defined as $\int_0^t g(t-s)\sigma(s)\mathrm{d}W_s$ for $t \in \mathbb{I}$, where σ is $(\mathcal{F}_t)_{t\in\mathbb{I}}$-predictable with locally bounded trajectories and finite second moments, and $g : \mathbb{I}\setminus\{0\} \to \mathbb{I}$ is Borel-measurable and square-integrable. If $\sigma \in \mathcal{C}^1(\mathbb{I})$, this class falls within the GFO framework.

Example 2.8 Bennedsen et al. [11] added a Gamma kernel to the volatility process, which yields the truncated Brownian semistationary (Bergomi-type) model

$$V_t = \xi_0(t)\mathcal{E}\left(2\nu C_H \int_0^t (t-s)^\alpha e^{-\beta(t-s)} \mathrm{d}W_s\right),$$

with $\beta > 0$, $\alpha \in (-\frac{1}{2}, \frac{1}{2})$. This corresponds to (2.3) with $Y = W$, Gamma fractional kernel $g(u) = u^\alpha e^{-\beta u}$ in (2.2) and

$$\Phi(\varphi)(t) := \xi_0(t)\exp\left(2\nu C_H \varphi(t)\right)\exp\left(-2\nu^2 C_H^2 \int_0^t (t-s)^{2\alpha} e^{-2\beta(t-s)} \mathrm{d}s\right).$$

Example 2.9 The version of the rough Heston model introduced by Guennoun et al. [33] reads

$$Y_t = Y_0 + \int_0^t \kappa(\theta - Y_s)\mathrm{d}s + \int_0^t \xi\sqrt{Y_s}\,\mathrm{d}W_s,$$

$$V_t = \eta + \int_0^t (t-s)^\alpha \mathrm{d}Y_s,$$

with Y_0, κ, ξ, θ all > 0, $2\kappa\theta > \xi^2$ and $\eta > 0$, $\alpha \in (-\frac{1}{2}, \frac{1}{2})$. This corresponds exactly to (2.3) with $g(u) = u^\alpha$, $\Phi(\varphi)(t) := \eta + \varphi(t)$, and the coefficients of (2.4) read $b(y) = \kappa(\theta - y)$ and $a(y) = \xi\sqrt{y}$. This model is markedly different from the rough Heston model introduced by El Euch and Rosenbaum [24] (for which the characteristic function is known in semiclosed form). Unfortunately, this version of the rough Heston model is outside of the scope of our invariance principle.

2.3 The approximation scheme

We now move on to the core of the project, namely an approximation scheme for the system (2.3), (2.4). The basic ingredient to construct approximating sequences will be suitable families of i.i.d. random variables which satisfy the following assumption.

Assumption 2.10 The family $(\xi_i)_{i\geq 1}$ forms an i.i.d. sequence of centered random variables with finite moments of all orders and $\mathbb{E}[\xi_1^2] = \sigma^2 > 0$.

Given $(\zeta_i)_{i\geq 1}$ as in Assumption 2.10, Lamperti's [49] generalisation of Donsker's invariance principle in Donsker [21] tells us that a Brownian motion W can be approximated weakly in Hölder space (Theorem 3.1) by processes of the form

$$W_n(t) := \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \zeta_k + \frac{nt - \lfloor nt \rfloor}{\sigma\sqrt{n}} \zeta_{\lfloor nt \rfloor + 1}, \tag{2.6}$$

defined pathwise for any $\omega \in \Omega$, $n \geq 1$ and $t \in \mathbb{I}$. As we show in Sect. 3.2, a similar construction holds to weakly approximate the process Y from (2.4) in Hölder space via

$$Y_n(t) := Y_n(0) + \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor} b(Y_n^{k-1}) + \frac{nt - \lfloor nt \rfloor}{n} b(Y_n^{\lfloor nt \rfloor})$$

$$+ \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} a(Y_n^{k-1})\zeta_k + \frac{nt - \lfloor nt \rfloor}{\sigma\sqrt{n}} a(Y_n^{\lfloor nt \rfloor})\zeta_{\lfloor nt \rfloor + 1}, \tag{2.7}$$

where $Y_n^k := Y_n(t_k)$ and $\mathcal{T}_n := \{t_k = \frac{k}{n} : k = 0, 1, \ldots, n\}$. Here the ζ_k correspond to the innovations of the Brownian motion W in (2.4). Similarly, we use ξ_k when referring to the innovations of the Brownian B from (2.3) which enter into the approximations of the log-stock price in (2.8) below. **Throughout the paper**,

we assume that the innovations $(\xi_i)_{i=1,\ldots,\lfloor nt \rfloor}$ and $(\zeta_i)_{i=1,\ldots,\lfloor nt \rfloor}$ come from two sequences $(\xi_i)_{i\geq 1}$ and $(\zeta_i)_{i\geq 1}$ satisfying Assumption 2.10 and such that $((\xi_i, \zeta_i))_{i\geq 1}$ is i.i.d. with $\mathrm{corr}(\xi_i, \zeta_i) = \rho$ for all $i \geq 1$. Naturally, the approximations in (2.7) and in (2.8) below should be understood pathwise, but we omit the ω-dependence in the notations for clarity.

Regarding the approximation scheme for the process X given by (2.3), we follow a typical route in weak convergence analysis (see Billingsley [13, Sect. 13] and Ethier and Kurtz [25, Chap. 3]) and establish convergence in the Skorokhod space $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$. Here $\mathcal{D}(\mathbb{I}) = \mathcal{D}(\mathbb{I}; \mathbb{R})$ denotes the space of $\mathbb{R}$-valued càdlàg functions on $\mathbb{I}$ and $d_{\mathcal{D}}$ denotes a metric inducing Skorokhod's J_1-topology. To approximate X in this space, we then consider the process

$$X_n(t) \quad := \quad -\frac{1}{2n} \sum_{k=1}^{\lfloor nt \rfloor} \Phi(\mathcal{G}^\alpha Y_n)(t_{k-1}) + \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \sqrt{\Phi(\mathcal{G}^\alpha Y_n)(t_{k-1})}\, \xi_k. \quad (2.8)$$

Analogously to (2.7), one could rewrite these as continuous processes via linear interpolation; but we note that the interpolating term would decay to zero by Chebyshev's inequality. The following result, proved in Sect. 3.4, confirms the functional convergence of the approximating sequence $(X_n)_{n\geq 1}$.

Theorem 2.11 *The sequence $(X_n)_{n\geq 1}$ converges in $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$ weakly to X.*

The construction of the proof allows extending the convergence to the case where Y is a d-dimensional diffusion without additional work. The proof of the theorem requires a certain number of steps: we start with the convergence of the approximations (Y_n) in some Hölder space, which we then translate into convergence of the sequence $(\Phi(\mathcal{G}^\alpha Y_n))$ by suitable continuity properties of the operations $\mathcal{G}^\alpha$ and Φ, before finally deducing also the convergence of the corresponding stochastic integrals for the approximations of (2.3). These steps are carried out in Sects. 3.2–3.4 below.

3 Functional CLT for a family of Hölder-continuous processes

3.1 Weak convergence of Brownian motion in Hölder spaces

Donsker's classical convergence result was proved under the Skorokhod topology. We concentrate here on convergence in the Hölder topology, due to Lamperti [49]. The standard convergence result for Brownian motion can be stated as follows.

Theorem 3.1 *For $\lambda < \frac{1}{2}$, the sequence (W_n) in (2.6) converges in $(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda)$ weakly to a Brownian motion.*

The proof relies on finite-dimensional convergence and tightness of the approximating sequence. Not surprisingly, the tightness criterion in [13, Sect. 13] for the Skorokhod space $\mathcal{D}(\mathbb{I})$ and for a Hölder space are different. In fact, the tightness criterion in a Hölder space is strictly related to the Kolmogorov–Čentsov continuity

result [16]. Note in passing that the approximating sequence (2.6) is piecewise differentiable in time for each $n \geq 1$ even though its limit is obviously not. The proof of Theorem 3.1 follows from Theorem 3.3 below under Assumption 2.10.

Theorem 3.2 *Let $Z \in \mathcal{C}^\lambda(\mathbb{I})$ and $(Z_n)_{n \geq 1}$ be an approximating sequence in the sense that for any sequence $(\tau_k)_{k \geq 1}$ in $\mathbb{I}$, the sequence $((Z_n(\tau_k))_{k \geq 1})_{n \geq 1}$ converges in distribution to $(Z(\tau_k))_{k \geq 1}$ as $n \to \infty$. Assume further that*

$$\mathbb{E}[|Z_n(t) - Z_n(s)|^\gamma] \leq C|t - s|^{1+\beta} \tag{3.1}$$

holds for all $n \geq 1$, $t, s \in \mathbb{I}$ and some C, γ, β all > 0 with $\frac{\beta}{\gamma} \leq \lambda$. Then $(Z_n)_{n \geq 1}$ converges in $\mathcal{C}^\mu(\mathbb{I})$ weakly to Z for $\mu < \frac{\beta}{\gamma} \leq \lambda$.

The proof of this theorem relies on results of Račkauskas and Suquet [63], who prove the convergence in the Hölder space $\mathcal{C}_0^\lambda(\mathbb{I})$ with the norm $\|f\|_\lambda^0 := |f|_\lambda + |f(0)|$ for all functions that satisfy

$$\lim_{\delta \downarrow 0} \sup_{\substack{0 < t - s < \delta \\ t, s \in \mathbb{I}}} \frac{|f(t) - f(s)|}{(t - s)^\gamma} = 0.$$

From this point, the proof of Theorem 3.2 is a straightforward consequence because $(\mathcal{C}_0^\lambda(\mathbb{I}), \|\cdot\|_\lambda^0)$ is a separable closed subspace of $(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda)$ (see Hamadouche [35] and [63] for details), and one can then use (3.1) to establish tightness as in [49]. Moreover, as the identity map from $C_0^\lambda(\mathbb{I})$ into $\mathcal{C}^\lambda(\mathbb{I})$ is continuous, weak convergence in the former implies weak convergence in the latter. To conclude our review of weak convergence in Hölder spaces, the following theorem due to Račkauskas and Suquet [63] provides necessary and sufficient conditions ensuring convergence in Hölder space.

Theorem 3.3 (Račkauskas–Suquet [63, Theorem 1]) *For $\lambda \in (0, \frac{1}{2})$, the sequence $(W_n)_{n \geq 1}$ in (2.6) converges in $\mathcal{C}^\lambda(\mathbb{I})$ weakly to a Brownian motion if and only if $\mathbb{E}[\xi_1] = 0$ and $\lim_{t \uparrow \infty} t^{\frac{1}{1 - 2\lambda}} \mathbb{P}[|\xi_1| \geq t] = 0$.*

Assumption 2.10 ensures the conditions in Theorem 3.3. The following statement allows us to apply Theorem 3.2 on $\mathbb{I}$ and extend the Hölder convergence result via linear interpolation to a sequence of continuous processes.

Theorem 3.4 *Let $Z \in \mathcal{C}^\lambda(\mathbb{I})$ and $(Z_n)_{n \geq 1}$ be an approximation sequence such that finite-dimensional convergence holds as $n \to \infty$. Moreover, if*

$$\mathbb{E}[|Z_n(t_i) - Z_n(t_j)|^\gamma] \leq C|t_i - t_j|^{1+\beta} \tag{3.2}$$

for any $t_i, t_j \in \mathcal{T}_n$ and some β, γ, C all > 0 with $\frac{\beta}{\gamma} \leq \lambda$ and $\gamma \geq 1 + \beta$, then the linearly interpolating sequence

$$\overline{Z}_n(t) := Z_n\left(\frac{\lfloor nt \rfloor}{n}\right) + (nt - \lfloor nt \rfloor)\left(Z_n\left(\frac{\lfloor nt \rfloor + 1}{n}\right) - Z_n\left(\frac{\lfloor nt \rfloor}{n}\right)\right)$$

satisfies (3.1). In particular, $(\overline{Z}_n)_{n \geq 1}$ converges in $\mathcal{C}^\mu(\mathbb{I})$ weakly to Z for $\mu < \frac{\beta}{\gamma} \leq \lambda$.

Proof It suffices to confirm (3.1) for $\bar{Z}_n$. The final claim then follows from Theorem 3.2. Take $t, s \in \mathbb{I}$ and suppose first that $t - s \geq \frac{1}{n}$. Letting $Z_n^k := Z_n(t_k)$ and $\overline{Z}_n^k := \overline{Z}_n(t_k)$, we can then write

$$
\begin{aligned}
\mathbb{E}[|\overline{Z}_n(t) - \overline{Z}_n(s)|^\gamma] &= \mathbb{E}[|Z_n^{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)(Z_n^{\lfloor nt \rfloor + 1} - Z_n^{\lfloor nt \rfloor}) \\
&\quad - Z_n^{\lfloor ns \rfloor} - (ns - \lfloor ns \rfloor)(Z_n^{\lfloor ns \rfloor + 1} - Z_n^{\lfloor ns \rfloor})|^\gamma] \\
&\leq 3^{\gamma - 1} \mathbb{E}[|Z_n^{\lfloor nt \rfloor} - Z_n^{\lfloor ns \rfloor}|^\gamma + [nt - \lfloor nt \rfloor]^\gamma |Z_n^{\lfloor nt \rfloor + 1} - Z_n^{\lfloor nt \rfloor}|^\gamma \\
&\quad + [ns - \lfloor ns \rfloor]^\gamma |Z_n^{\lfloor ns \rfloor + 1} - Z_n^{\lfloor ns \rfloor}|^\gamma] \\
&\leq C\left(\left(\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n}\right)^{1+\beta} + \frac{(nt - \lfloor nt \rfloor)^\gamma}{n^{1+\beta}} + \frac{(ns - \lfloor ns \rfloor)^\gamma}{n^{1+\beta}}\right) \\
&\leq C(t - s)^{1+\beta},
\end{aligned}
$$

where we used (3.2) and the fact that $\frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \leq 2(t - s)$, $nt - \lfloor nt \rfloor \leq 1$ for $t \geq 0$ since $t - s \geq \frac{1}{n}$.

It remains to consider the case $t - s < \frac{1}{n}$. There are two possible scenarios. If $\lfloor nt \rfloor = \lfloor ns \rfloor$, then using $\gamma \geq 1 + \beta$ gives

$$
\mathbb{E}[|\overline{Z}_n(t) - \overline{Z}_n(s)|^\gamma] = \mathbb{E}[|(nt - ns)(Z_n^{\lfloor nt \rfloor + 1} - Z_n^{\lfloor nt \rfloor})|^\gamma] \leq \frac{C|t - s|^\gamma}{n^{1+\beta-\gamma}} \leq C(t-s)^{1+\beta}.
$$

If $\lfloor nt \rfloor \neq \lfloor ns \rfloor$, then either $\lfloor nt \rfloor + 1 = \lfloor ns \rfloor$ or $\lfloor nt \rfloor = \lfloor ns \rfloor + 1$. Without loss of generality, consider the second case. Then

$$
\begin{aligned}
\mathbb{E}[|\overline{Z}_n(t) - \overline{Z}_n(s)|^\gamma] &= \mathbb{E}[|\overline{Z}_n(t) - Z_n^{\lfloor nt \rfloor} + Z_n^{\lfloor nt \rfloor} - \overline{Z}_n(s)|^\gamma] \\
&\leq 2^{\gamma - 1} \mathbb{E}[|\overline{Z}_n(t) - Z_n^{\lfloor nt \rfloor}|^\gamma + |Z_n^{\lfloor nt \rfloor} - \overline{Z}_n(s)|^\gamma] \\
&\leq C\left((t - s)^{1+\beta} + \mathbb{E}[|(\lfloor nt \rfloor - ns)(Z_n^{\lfloor nt \rfloor} - Z_n^{\lfloor nt \rfloor - 1})|^\gamma]\right),
\end{aligned}
$$

and the result follows as before since $t - \frac{\lfloor nt \rfloor}{n} < |t - s|$ and $|s - \frac{\lfloor nt \rfloor}{n}| \leq |t - s|$. $\qquad \square$

3.2 Weak convergence of Itô diffusions in Hölder spaces

The first important step in our analysis is to extend the Donsker–Lamperti weak convergence from Brownian motion to the Itô diffusion Y in (2.4).

Theorem 3.5 *The sequence* $(Y_n)_{n \geq 1}$ *in (2.7) converges in* $(\mathcal{C}^\lambda(\mathbb{I}), \| \cdot \|_\lambda)$ *weakly to* Y *in (2.4) for all* $\lambda < \frac{1}{2}$,

Proof Finite-dimensional convergence is a classical result by Kushner [47]; so only tightness needs to be checked. In particular, using Theorem 3.4, we need only con-

sider the partition $\mathcal{T}_n$. Thus we get

$$\mathbb{E}[|Y_n^j - Y_n^i|^{2p}]$$

$$= \mathbb{E}\left[\left|\sum_{k=i+1}^{j} \frac{1}{n}b(Y_n^{k-1}) + \frac{1}{\sigma\sqrt{n}}a(Y_n^{k-1})\zeta_k\right|^{2p}\right]$$

$$\leq 2^{2p-1}\left(\mathbb{E}\left[\left|\sum_{k=i+1}^{j} \frac{1}{n}b(Y_n^{k-1})\right|^{2p}\right] + \mathbb{E}\left[\left|\sum_{k=i+1}^{j} \frac{a(Y_n^{k-1})\zeta_k}{\sigma\sqrt{n}}\right|^{2p}\right]\right)$$

$$\leq 2^{2p-1}\left(\mathbb{E}\left[\left|\sum_{k=i+1}^{j} \frac{1}{n}b(Y_n^{k-1})\right|^{2p}\right] + C(p)\mathbb{E}\left[\left|\sum_{k=i+1}^{j} \frac{a(Y_n^{k-1})^2\zeta_k^2}{\sigma^2 n}\right|^{p}\right]\right)$$

$$\leq 2^{2p-1}\left(\frac{(j-i)^{2p-1}}{n^{2p}}\sum_{k=i+1}^{j}\mathbb{E}[b(Y_n^{k-1})^{2p}]\right.$$

$$\left. + \frac{(j-i)^{p-1}}{n^p}C(p)\frac{\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\sum_{k=i+1}^{j}\mathbb{E}[a(Y_n^{k-1})^{2p}]\right)$$

$$\leq 2^{2p-1}\frac{(j-i)^{p-1}}{n^p}\sum_{k=i+1}^{j}\left(C_b^{2p}\mathbb{E}[(1+|Y_n^{k-1}|)^{2p}]\right.$$

$$\left. + C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\mathbb{E}[(1+|Y_n^{k-1}|)^{2p}]\right)$$

$$\leq \max\left(C_b^{2p}, C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\right)2^{2p}\frac{(j-i)^{p-1}}{n^p}\left((j-i) + \sum_{k=i+1}^{j}\mathbb{E}[|Y_n^{k-1}|^{2p}]\right)$$

$$\leq \max\left(C_b^{2p}, C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\right)2^{2p}\exp\left(\sum_{k=i+1}^{j}2^{2p}\frac{(j-i)^{p-1}}{n^p}\right)(t_j - t_i)^p$$

$$\leq \max\left(C_b^{2p}, C(p)\frac{C_a^{2p}\mathbb{E}[\zeta_1^{2p}]}{\sigma^{2p}}\right)2^{2p}\exp(2^{2p})(t_j - t_i)^p$$

$$=: \mathfrak{C}(p)(t_j - t_i)^p,$$

where we have used the discrete version of the BDG inequality (see Beiglböck and Siorpaes [8, Theorem 6.3]) in the martingale term $\sum_{k=i+1}^{j}\frac{1}{\sigma\sqrt{n}}a(Y_n^{k-1})\zeta_k$ with the constant $C(p) := 6^p(p-1)^{p-1}$. Indeed, if we consider the discrete-time martingale process $(x_n^{i,j})_u := \sum_{k=1}^{u}\frac{1}{\sigma\sqrt{n}}a(Y_n^{k+i-1})\zeta_{i+k}$ for $u \in \{1, \ldots, j-i\}$, we have $|(x_n^{i,j})_{j-i}| \leq \max_{u \in \{1,\ldots,j-i\}}|(x_n^{i,j})_u|$, and the BDG inequality clearly also applies to $|x_{j-i}^{i,j}|$. We also used independence of ζ_k and Y_{k-1} and the linear growth of $b(\cdot)$ and $a(\cdot)$ from Assumption 2.3, Hölder's inequality and the discrete version of Gronwall's lemma (see Clark [17]) in the last step. Since $\mathbb{E}[\zeta_k^{2p}]$ is bounded by Assumption 2.10 and the constant $\mathfrak{C}(p)$ only depends on p, but not on n, the criterion (3.2) of Theorem 3.4 holds for $p > 1$ with $\gamma = 2p$ and $\beta = p - 1$. $\qquad\square$

Corollary 3.6 *Let $(Y_n)_{n\geq 1}$ be as in Theorem 3.5 with innovations $(\zeta_i)_{i\geq 1}$, and suppose $(B^n)_{n\geq 1}$ is defined by the Donsker approximations (2.6) for some innovations $(\xi_i)_{i\geq 1}$ satisfying Assumption 2.10 such that $((\zeta_i, \xi_i))_{i\geq 1}$ is i.i.d. with $\mathrm{corr}(\zeta_i, \xi_i) = \rho$ for all $i \geq 1$. Then there is joint weak convergence in $(\mathcal{C}^\lambda(\mathbb{I}, \mathbb{R}^2), \|\cdot\|_\lambda)$ of (B_n, Y_n) to (B, Y), for all $\lambda < \frac{1}{2}$, for a standard Brownian motion B such that $[B, W]_t = \rho t$ for $t \in \mathbb{I}$, where W is the standard Brownian motion driving the dynamics of the weak limit Y in (2.4).*

Proof Let $(\zeta_i^\perp)_{i\geq 1}$ satisfy Assumption 2.10 and be independent of the innovations $(\zeta_i)_{i\geq 1}$ defining $(Y^n)_{n\geq 1}$. Then set $\xi_i := \rho\zeta_i + \overline{\rho}\zeta_i^\perp$ for $i \geq 1$ and let B_n be defined in terms of $(\xi_i)_{i\geq 1}$. This yields the same finite-dimensional distributions of (B_n, Y_n) as for the general $(\xi_i)_{i\geq 1}$ in the statement of Corollary 3.6. Consider now the drift vector $\mathbf{b}(y) = (0, b(y))$ and the 2×2 diffusion matrix $\mathbf{a}(y)$ with rows $(\rho, \overline{\rho})$ and $(0, a(y))$. Then the result of Kushner [47] applies directly to give finite-dimensional convergence with the desired limit. Finally, tightness of $((B_n, Y_n))_{n\geq 1}$ follows analogously to the proof of Theorem 3.5, and the claim follows. □

3.3 Invariance principle for rough processes

We have set the ground to extend our results to processes that are not necessarily $(\frac{1}{2} - \varepsilon)$-Hölder-continuous, Markovian or semimartingales. More precisely, we are interested in α-Hölder-continuous paths with $\alpha \in (0, 1)$, such as Riemann–Liouville fractional Brownian motion or some TBSS processes. A key tool is the continuous mapping theorem, first proved by Mann and Wald [53], which establishes the preservation of weak convergence under continuous operators.

Theorem 3.7 *Let $(\mathcal{X}, \|\cdot\|_\mathcal{X})$ and $(\mathcal{Y}, \|\cdot\|_\mathcal{Y})$ be two normed spaces and assume that $g : \mathcal{X} \to \mathcal{Y}$ is a continuous operator. If the sequence $(Z_n)_{n\geq 1}$ of random variables converges in $(\mathcal{X}, \|\cdot\|_\mathcal{X})$ weakly to Z, then $(g(Z_n))_{n\geq 1}$ converges in $(\mathcal{Y}, \|\cdot\|_\mathcal{Y})$ weakly to $g(Z)$.*

Many authors have exploited the combination of Theorems 3.1 and 3.7 to prove weak convergence; see e.g. Pollard [62, Chap. IV]. This path avoids the lengthy computations of tightness and finite-dimensional convergence in classical proofs. In fact, Hamadouche [35] already realised that Riemann–Liouville fractional operators are continuous, so that Theorem 3.7 holds under mapping by Hölder-continuous functions. In contrast, the novelty here is to consider the family of GFOs applied to Brownian motion together with the extension of Brownian motion to Itô diffusions.

The analogue of Theorem 3.5 for $\mathcal{G}^\alpha Y$ reads as follows.

Theorem 3.8 *For $(Y_n)_{n\geq 1}$ in (2.7) and its weak limit Y in $(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda)$ for $\lambda < \frac{1}{2}$, the representation*

$$(\mathcal{G}^\alpha Y_n)(t) = \sum_{i=1}^{\lfloor nt \rfloor} n\big(G(t - t_{i-1}) - G(t - t_i)\big)(Y_n^i - Y_n^{i-1})$$

$$+ nG(t - t_{\lfloor nt \rfloor})\big(Y_n(t) - Y_n^{\lfloor nt \rfloor}\big) \tag{3.3}$$

holds for $t \in \mathbb{I}$. Furthermore, this sequence $(\mathcal{G}^\alpha Y_n)_{n \geq 1}$ converges in $(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda})$ weakly to $\mathcal{G}^\alpha Y$ for any $\alpha \in (-\lambda, 1 - \lambda)$.

Proof Recall that the sequence (2.7) is piecewise differentiable in time. First, for $\alpha \in (0, 1 - \lambda)$, note that $g(0) = 0$ since $\alpha > 0$ and therefore by integration by parts (where Y_n is piecewise differentiable), for $n \geq 1$ and $t \in \mathbb{I}$,

$$(\mathcal{G}^\alpha Y_n)(t)$$

$$= \int_0^t g'(t - s)\big(Y_n(s) - Y_n(0)\big)ds$$

$$= \int_0^t g(t - s)\frac{d(Y_n(s) - Y_n(0))}{ds}ds$$

$$= \frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^{\lfloor nt \rfloor} n\int_{t_{i-1}}^{t_i} g(t - s)a(Y_n^{i-1})\zeta_i ds + n\int_{t_{\lfloor nt \rfloor}}^t g(t - s)a(Y_n^{\lfloor nt \rfloor})\zeta_{\lfloor nt \rfloor+1}ds\right)$$

$$+ \frac{1}{n}\left(n\sum_{i=1}^{\lfloor nt \rfloor}\int_{t_{i-1}}^{t_i} g(t - s)b(Y_n^{i-1})ds + n\int_{t_{\lfloor nt \rfloor}}^t g(t - s)b(Y_n^{\lfloor nt \rfloor})ds\right)$$

$$= \sum_{i=1}^{\lfloor nt \rfloor} n\big(G(t - t_{i-1}) - G(t - t_i)\big)(Y_n^i - Y_n^{i-1})$$

$$+ n\big(G(t - t_{\lfloor nt \rfloor}) - G(0)\big)\big(Y_n(t) - Y_n^{\lfloor nt \rfloor}\big),$$

and (3.3) follows since $G(0) = 0$ in the last line. When $\alpha \in (-\lambda, 0)$, using $G(0) = 0$, we similarly get

$$\int_0^t g(t - s)\big(Y_n(s) - Y_n(0)\big)ds$$

$$= \int_0^t G(t - s)\frac{d(Y_n(s) - Y_n(0))}{ds}ds$$

$$= \frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^{\lfloor nt \rfloor} n\int_{t_{i-1}}^{t_i} G(t - s)a(Y_n^{i-1})\zeta_i ds + n\int_{t_{\lfloor nt \rfloor}}^t G(t - s)a(Y_n^{\lfloor nt \rfloor})\zeta_{\lfloor nt \rfloor+1}ds\right)$$

$$+ \frac{1}{n}\left(n\sum_{i=1}^{\lfloor nt \rfloor}\int_{t_{i-1}}^{t_i} G(t - s)b(Y_n^{i-1})ds + n\int_{t_{\lfloor nt \rfloor}}^t G(t - s)b(Y_n^{\lfloor nt \rfloor})ds\right)$$

$$= n\left(\sum_{i=1}^{\lfloor nt \rfloor}\left(\frac{b(Y_n^{i-1})}{n} + \frac{a(Y_n^{i-1})}{\sigma\sqrt{n}}\zeta_i\right)\int_{t_{i-1}}^{t_i} G(t - s)ds\right.$$

$$+ \left.\left(\frac{b(Y_n^{\lfloor nt \rfloor})}{n} + \frac{a(Y_n^{\lfloor nt \rfloor})}{\sigma\sqrt{n}}\zeta_{\lfloor nt \rfloor+1}\right)\int_{t_{\lfloor nt \rfloor}}^t G(t - s)ds\right)$$

$$= n\left(\sum_{i=1}^{\lfloor nt \rfloor}(Y_n^i - Y_n^{i-1})\int_{t_{i-1}}^{t_i} G(t - s)ds + \big(Y_n(t) - Y_n^{\lfloor nt \rfloor}\big)\int_{t_{\lfloor nt \rfloor}}^t G(t - s)ds\right),$$

and from there it readily follows that

$$
\begin{aligned}
(\mathcal{G}^\alpha Y_n)(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t g(t-s)\big(Y_n(s)-Y_n(0)\big)\mathrm{d}s \\
&= \sum_{i=1}^{\lfloor nt \rfloor} n\big(G(t-t_{i-1})-G(t-t_i)\big)(Y_n^i - Y_n^{i-1}) \\
&\quad + nG(t-t_{\lfloor nt \rfloor})\big(Y_n(t)-Y_n^{\lfloor nt \rfloor}\big)
\end{aligned}
$$

as desired; note that when $t = \frac{k}{n}$, the difference quotients pick up an extra term, but this vanishes in the limit. Finally, the claimed convergence follows analogously to that in Theorem 3.5 by continuous mapping, along with the fact that $\mathcal{G}^\alpha$ is a continuous operator from $(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda)$ to $(\mathcal{C}^{\lambda+\alpha}(\mathbb{I}), \|\cdot\|_{\lambda+\alpha})$ for all $\lambda \in (0,1)$ and $\alpha \in (-\lambda, 1-\lambda)$. □

Notice here that the mean value theorem implies

$$
(\mathcal{G}^\alpha Y_n)(t) = \sum_{i=1}^{\lfloor nt \rfloor} g(t_i^*)(Y_n^i - Y_n^{i-1}) + g(t_{\lfloor nt \rfloor+1}^*)\big(Y_n(t)-Y_n^{\lfloor nt \rfloor}\big), \qquad (3.4)
$$

where $t_i^* \in [t-t_i, t-t_{i-1}]$ and $t_{\lfloor nt \rfloor+1}^* \in [0, t-t_{\lfloor nt \rfloor}]$ and we use that $G(0)=0$. This expression is closer to the usual left-point forward Euler approximation. For numerical purposes, (3.4) is much more efficient since the integral G required in (3.3) is not necessarily available in closed form. Nevertheless, not any arbitrary choice of t_i^* gives the desired convergence from the above argument. We present a suitable candidate for optimal t_i^* in Sect. 4.3.1, which guarantees weak convergence in the Hölder sense.

As could be expected, the Hurst parameter influences the speed of convergence of the scheme. We leave a formal proof to further study, but the following argument provides some intuition about the correct normalising factor. Given $g \in \mathcal{L}^\alpha$, we can write $g(u) = u^\alpha L(u)$, where L is a bounded function on $\mathbb{I}$. At time $t = t_i$, take $t_k^* = t_i - t_k + \frac{\varepsilon}{n}$ for $\varepsilon \in [0,1]$. For $\alpha \in (-\lambda, 0)$, since $g \in \mathcal{L}^\alpha$, we can rewrite the approximation (3.4) as

$$
(\mathcal{G}^\alpha Y_n)(t_i) = \frac{1}{n^{1/2+\alpha}}\sum_{k=1}^{i}(i-k+\varepsilon)^\alpha L(t_k^*)(Y_n^k - Y_n^{k-1})\sqrt{n} \qquad \text{for } i = 0,\ldots,n.
$$

Here, $(i-k+\varepsilon)^\alpha \le \varepsilon^\alpha$ is bounded in $n \ge 1$ as long as $\varepsilon \in (0,1]$; so the normalisation factor is of the order $n^{-\alpha-1/2}$. When $\alpha \in (0, 1-\lambda)$, the approximation (3.4) instead reads
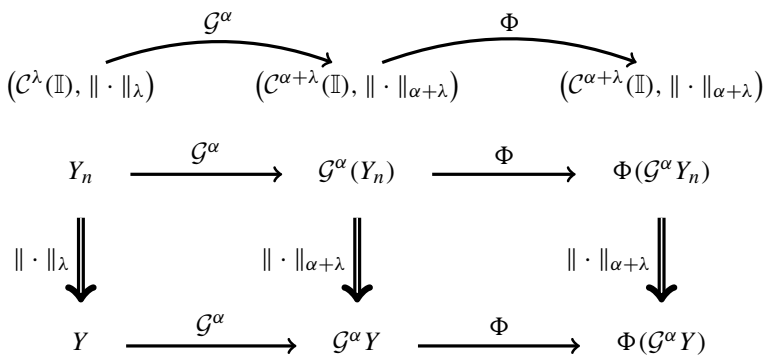
$$
(\mathcal{G}^\alpha Y_n)(t_i) = \frac{1}{\sqrt{n}}\sum_{k=1}^{i}\left(t_i - t_k + \frac{\varepsilon}{n}\right)^\alpha L(t_k^*)(Y_n^k - Y_n^{k-1})\sqrt{n} \qquad \text{for } i = 0,\ldots,n,
$$

in which case $(t_i - t_k + \frac{\varepsilon}{n})^\alpha \le t_i^\alpha$ is bounded in $n \ge 1$, and hence the normalisation factor is of the order $n^{-1/2}$. This intuition is consistent with the result by Neuenkirch

and Shalaiko [57] who found the strong rate of convergence of the Euler scheme to be of the order $\mathcal{O}(n^{-H})$ for $H < \frac{1}{2}$ for fractional Ornstein–Uhlenbeck. So far, our results hold for α-Hölder-continuous functions; however, for practical purposes, it is often necessary to constrain the volatility process $(V_t)_{t\in\mathbb{I}}$ to remain strictly positive at all times. The stochastic integral $\mathcal{G}^\alpha Y$ need not be such in general. However, a simple transformation (e.g. exponential) can easily overcome this fact. The remaining question is whether the α-Hölder-continuity is preserved after this composition.

Proposition 3.9 *Let $(Y_n)_{n\geq1}$ be the approximating sequence (2.7) in $\mathcal{C}^\lambda(\mathbb{I})$ for fixed $\lambda < 1/2$. Then $(\Phi(\mathcal{G}^\alpha Y_n))_{n\geq1}$ converges in $(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda})$ weakly to $\Phi(\mathcal{G}^\alpha Y)$ for all $\alpha \in (-\lambda, 1-\lambda)$.*

Proof By Theorem 3.8, $(\mathcal{G}^\alpha Y_n)_{n\geq1}$ converges in $(\mathcal{C}^{\lambda+\alpha}(\mathbb{I}), \|\cdot\|_{\lambda+\alpha})$ weakly to $\mathcal{G}^\alpha Y$. Furthermore, with our assumptions, Φ is continuous from $(\mathcal{C}^{\lambda+\alpha}(\mathbb{I}), \|\cdot\|_{\lambda+\alpha})$ to $(\mathcal{C}^{\lambda+\alpha}(\mathbb{I}), \|\cdot\|_{\lambda+\alpha})$. The proposition thus follows from the continuous mapping theorem. The diagram below summarises the steps with $\lambda < 1/2$. The double arrows show weak convergence, and we indicate next to them the topology in which it takes place.

$$
\begin{array}{ccccc}
\big(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda\big) & \xrightarrow{\mathcal{G}^\alpha} & \big(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda}\big) & \xrightarrow{\Phi} & \big(\mathcal{C}^{\alpha+\lambda}(\mathbb{I}), \|\cdot\|_{\alpha+\lambda}\big) \\[2mm]
Y_n & \xrightarrow{\mathcal{G}^\alpha} & \mathcal{G}^\alpha(Y_n) & \xrightarrow{\Phi} & \Phi(\mathcal{G}^\alpha Y_n) \\[1mm]
\|\cdot\|_\lambda \Downarrow & & \|\cdot\|_{\alpha+\lambda} \Downarrow & & \|\cdot\|_{\alpha+\lambda} \Downarrow \\[1mm]
Y & \xrightarrow{\mathcal{G}^\alpha} & \mathcal{G}^\alpha Y & \xrightarrow{\Phi} & \Phi(\mathcal{G}^\alpha Y)
\end{array}
$$
$\square$

3.4 Extending the weak convergence to the Skorokhod space and proof of Theorem 2.11

The Skorokhod space $(\mathcal{D}(\mathbb{I}), d_\mathcal{D})$ of càdlàg functions equipped with the Skorokhod topology has been widely used to prove weak convergence. It markedly simplifies when we only consider continuous functions (as is the case for our framework with Hölder-continuous processes). Billingsley [13, Sect. 12] proved that the identity $(\mathcal{D}(\mathbb{I}) \cap \mathcal{C}(\mathbb{I}), d_\mathcal{D}) = (\mathcal{C}(\mathbb{I}), \|\cdot\|_\infty)$ always holds. This seemingly simple statement allows us to reduce proofs of weak convergence of continuous processes in the Skorokhod topology to that in the supremum norm, which is usually much simpler. We start with the following straightforward observation.

Lemma 3.10 *For $\lambda \in (0, 1)$, the identity map is continuous from $(\mathcal{C}^\lambda(\mathbb{I}), \|\cdot\|_\lambda)$ to $(\mathcal{D}(\mathbb{I}), d_\mathcal{D})$.*

Proof Since the identity map is linear, it suffices to check that it is bounded. For this observe that $\|f\|_\lambda = |f|_\lambda + \sup_{t \in \mathbb{I}} |f(t)| = |f|_\lambda + \|f\|_\infty > \|f\|_\infty$, which concludes the proof since the Skorokhod norm in the space of continuous functions is equivalent to the supremum norm. $\qquad\square$

Applying the continuous mapping theorem twice, first with the generalised fractional operator (Theorem 3.8) and then with the identity map, directly yields the following result.

Theorem 3.11 *For any $\alpha \in (-\frac{1}{2}, \frac{1}{2})$, the sequence $(\Phi(\mathcal{G}^\alpha Y_n))_{n \geq 1}$ converges in $(\mathcal{D}(\mathbb{I}), d_\mathcal{D})$ weakly to $\Phi(\mathcal{G}^\alpha Y)$. Moreover, the sequence is tight in $(\mathcal{C}(\mathbb{I}), \|\cdot\|_\infty)$.*

The final step in the proof of our main theorem is to extend the functional weak convergence to the log-stock price X. For this, we rely on the weak convergence theory for stochastic integrals due to Jakubowski et al. [42] and further developed by Kurtz and Protter [46]. **Throughout**, we write $H \bullet N := \int H \, dN$ and use the notation H_- for the process $H_-(t) := H(t-)$ obtained by taking left limits. The next result is a restatement of [46, Theorem 2.2] in the special case $\delta = \infty$ (in their notation) and restricted to real-valued processes on $\mathbb{I}$.

Theorem 3.12 *For each $n \geq 1$, let $N_n = M_n + A_n$ be an $(\mathcal{F}_t^n)$-semimartingale and H_n an $(\mathcal{F}_t^n)$-adapted càdlàg process on $\mathbb{I}$. Suppose that for all $\gamma > 0$, there are $(\mathcal{F}_t^n)$-stopping times $(\tau_n^\gamma)_{n \geq 1}$ with the property that $\sup_{n \geq 1} \mathbb{P}[\tau_n^\gamma \leq \gamma] \leq 1/\gamma$ and $\sup_{n \geq 1} \mathbb{E}[[M_n]_{\tau_n^\gamma \wedge 1} + T_{\tau_n^\gamma \wedge 1}(A_n)] < \infty$, where T_t denotes the total variation on $[0, t]$. If $(H_n, N_n)_{n \geq 1}$ converges in $(\mathcal{D}(\mathbb{I}, \mathbb{R}^2), d_\mathcal{D})$ weakly to (H, N), then N is a semimartingale in the filtration generated by (H, N) and $(H_n, N_n, (H_n)_- \bullet N_n)_{n \geq 1}$ converges in $(\mathcal{D}(\mathbb{I}, \mathbb{R}^3), d_\mathcal{D})$ weakly to $(H, N, H_- \bullet N)$.*

With this, we can now give the proof of Theorem 2.11 which asserts the functional weak convergence of the approximations $(X_n)_{n \geq 1}$ from (2.8) to the desired log-price X from (2.3).

Proof *of Theorem 2.11* We begin by considering, for all $n \geq 1$, the particular approximations

$$M_n(t) := \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \xi_i,$$

for $t \in \mathbb{I}$, of the driving Brownian motion B in the dynamics of X. Here the ξ_i satisfy Assumption 2.10, and so do the ζ_i in the construction of Y_n from (2.7). While each pair ξ_i and ζ_i are correlated, they form an i.i.d. sequence $((\zeta_i, \xi_i))_{i \geq 1}$ across the pairs. In particular, it is straightforward to see that each M_n is a martingale on $\mathbb{I}$ for the filtration $(\mathcal{F}_t^n)$ defined by $\mathcal{F}_t^n := \sigma(\zeta_i, \xi_i : i = 1, \ldots, \lfloor nt \rfloor)$. Moreover, we have

$$\mathbb{E}[[M_n]_t] = \lfloor nt \rfloor \frac{1}{\sigma^2 n} \mathbb{E}[\xi_1^2] = \frac{\lfloor nt \rfloor}{n} \leq 1$$

for $t \in \mathbb{I}$ and for all $n \geq 1$. Consequently, we can simply take $\tau_n^\gamma \equiv +\infty$ for all $\gamma > 0$ and $n \geq 1$ to satisfy the required control on the integrators $N_n := M_n$ in Theorem 3.12. By Ethier and Kurtz [25, Theorem 7.1.4], the M_n converge in $(\mathcal{D}(\mathbb{I}), d_\mathcal{D})$ weakly to a Brownian motion B. Now fix $\alpha \in (-\frac{1}{2}, \frac{1}{2})$ and define a sequence $(H_n)_{n \geq 1}$ of càdlàg processes on $\mathbb{I}$ by setting $H_n(1) := \Phi(\mathcal{G}^\alpha Y_n)(1)$ and $H_n(t) := \Phi(\mathcal{G}^\alpha Y_n)(t_{k-1})$ for $t \in [t_{k-1}, t_k)$ and $k = 1, \ldots, n$. In view of Theorem 3.11, the Arzelà–Ascoli characterisation of tightness (see [13, Theorem 8.2]) for the space $(\mathcal{C}(\mathbb{I}), \|\cdot\|_\infty)$ allows us to conclude that the H_n converge in $(\mathcal{D}(\mathbb{I}), d_\mathcal{D})$ weakly to $H := \Phi(\mathcal{G}^\alpha Y)$. Furthermore, recalling the definition of Y_n in (2.7), each H_n is adapted to the filtration $(\mathcal{F}_t^n)$ introduced above. By Corollary 3.6, we readily deduce that there is joint weak convergence in $(\mathcal{D}(\mathbb{I}), d_\mathcal{D}) \times (\mathcal{D}(\mathbb{I}), d_\mathcal{D}) \times (\mathcal{D}(\mathbb{I}), d_\mathcal{D})$ of (Y_n, H_n, M_n) to (Y, H, B), where Y satisfies (2.4) for a Brownian motion W with $[W, B]_t = \rho t$ for all $t \in \mathbb{I}$. As noted in [46], the Skorokhod topology on $\mathcal{D}(\mathbb{I}, \mathbb{R}^2)$ is stronger than the product topology on $\mathcal{D}(\mathbb{I}) \times \mathcal{D}(\mathbb{I})$, but here it automatically follows that we have weak convergence in $(\mathcal{D}(\mathbb{I}, \mathbb{R}^2), d_\mathcal{D})$ of the pairs (H_n, M_n) to (H, B) by standard properties of the Skorokhod topology (see e.g. [25, Theorem 3.10.2]), since the limiting pair (H, B) is continuous. Consequently, we are in a position to apply Theorem 3.12. To this end, observe that

$$\big((H_n)_- \bullet M_n\big)(t) = \sum_{k=1}^{\lfloor nt \rfloor} H_n(t_k-)\big(M_n(t_k) - M_n(t_k-)\big) = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \Phi(\mathcal{G}^\alpha Y_n)(t_{k-1}) \xi_k,$$

which is precisely the second term on the right-hand side of (2.8). Therefore, Theorem 3.12 gives that the stochastic integral $H \bullet M = \sqrt{\Phi(\mathcal{G}^\alpha Y)} \bullet B$ is in $(\mathcal{D}(\mathbb{I}), d_\mathcal{D})$ the weak limit of the second term on the right-hand side of (2.8). For the first term on the right-hand side of (2.8), we have $-\frac{1}{2} \int_0^{\cdot} H_n(s) \mathrm{d}s$ converging weakly to $-\frac{1}{2} \int_0^{\cdot} H(s) \mathrm{d}s$ by the continuous mapping theorem, as the integral is a continuous operator from $(\mathcal{D}(\mathbb{I}), d_\mathcal{D})$ to itself. Since there is weak convergence in $(\mathcal{D}(\mathbb{I}, \mathbb{R}^2), d_\mathcal{D})$ of the pairs $(H_n, (H_n)_- \bullet M_n)$ to $(H, H \bullet B)$, the sum of the two terms on the right-hand side of (2.8) are then also weakly convergent in $(\mathcal{D}(\mathbb{I}), d_\mathcal{D})$. Recalling that the limit Y satisfies (2.4) for a Brownian motion W such that W and B are correlated with parameter ρ, we hence conclude that X converges in $(\mathcal{D}(\mathbb{I}), d_\mathcal{D})$ weakly to the desired limit. □

4 Applications

4.1 Weak convergence of the hybrid scheme

The hybrid scheme (and its turbocharged version in McCrickerd and Pakkanen [54]) introduced by Bennedsen et al. [10] is the current state-of-the-art to simulate TBSS processes. However, only convergence in mean-square-error was proved, but not (functional) weak convergence which would justify the use of the scheme for path-dependent options. Unless otherwise stated, we denote by $\mathcal{T}_n := \{t_k = \frac{k}{n} : k = 0, 1, \ldots, n\}$ the uniform grid on $\mathbb{I}$. We show that the Hölder convergence also holds for the case $g(x) = x^\alpha$.

Proposition 4.1 *The hybrid scheme sequence $(\widetilde{\mathcal{G}}^\alpha W_n)_{n \geq 1}$ is defined as*

$$\widetilde{\mathcal{G}}^\alpha W_n(t) := \widetilde{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor}{n}\right) + (nt - \lfloor nt \rfloor)\left(\widetilde{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor + 1}{n}\right) - \widetilde{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor}{n}\right)\right)$$

for $t \in \mathbb{I}$, where for $i = 0, \ldots, n$ and $\kappa \geq 1$,

$$\widetilde{\mathcal{G}}^\alpha W_n(t_i) := \sum_{k=1}^{(i-\kappa) \vee 0} \xi_k \int_{t_{k-1}}^{t_k} \sqrt{n}(t_i - s)^\alpha \, \mathrm{d}s + \int_{0 \vee t_{i-\kappa}}^{t_i} (t_i - s)^\alpha \, \mathrm{d}W_s$$

with $\xi_k := \int_{t_{k-1}}^{t_k} \mathrm{d}W_s \sim \mathcal{N}(0, \frac{1}{n})$. It converges in $(\mathcal{C}^{\alpha+1/2}, \|\cdot\|_{\alpha+1/2})$ weakly to $\mathcal{G}^\alpha W$ for $\alpha \in (-\frac{1}{2}, \frac{1}{2})$ and $\kappa = 1$.

Proof Finite-dimensional convergence follows trivially as the target process is centered Gaussian; thus convergence of the covariance matrices ensures finite-dimensional convergence. To prove weak convergence, we only need to show that the approximating sequence is tight, by verifying the criteria from Theorem 3.4 as follows. In Theorem 3.4, we need to show that

$$\mathbb{E}[|\widetilde{\mathcal{G}}^\alpha W_n(t_i) - \widetilde{\mathcal{G}}^\alpha W_n(t_j)|^{2p}] \leq C|t_i - t_j|^{2p\alpha + p}$$

for all $t_i, t_j \in \mathcal{T}_n$, for $p \geq 1$ and some constant $C \geq 0$. Without loss of generality, assume $t_j < t_i$ and take $\kappa = 1$. Define

$$\widetilde{\sigma}^2 := \mathbb{E}[|\widetilde{\mathcal{G}}^\alpha W_n(t_i) - \widetilde{\mathcal{G}}^\alpha W_n(t_j)|^2].$$

The statement is trivial if $j = 0$ or $i = 1$, so we assume otherwise. By the definition of $\widetilde{\mathcal{G}}^\alpha W_n(t_i)$ and using the notation $\lesssim$ to mean less than or equal with a constant factor, we can write, since $\kappa = 1$,

$$|\widetilde{\mathcal{G}}^\alpha W_n(t_i) - \widetilde{\mathcal{G}}^\alpha W_n(t_j)|^2$$

$$= \left| \sqrt{n} \sum_{k=1}^{i-1} \xi_k \int_{t_{k-1}}^{t_k} (t_i - s)^\alpha \, \mathrm{d}s - \sqrt{n} \sum_{k=1}^{j-1} \xi_k \int_{t_{k-1}}^{t_k} (t_j - s)^\alpha \, \mathrm{d}s \right.$$
$$\left. + \int_{t_{i-1}}^{t_i} (t_i - s)^\alpha \, \mathrm{d}W_s - \int_{t_{j-1}}^{t_j} (t_j - s)^\alpha \, \mathrm{d}W_s \right|^2$$

$$= \left| \sqrt{n} \sum_{k=1}^{j-1} \xi_k \int_{t_{k-1}}^{t_k} \left((t_i - s)^\alpha - (t_j - s)^\alpha\right) \mathrm{d}s - \sqrt{n} \sum_{k=j}^{i-1} \xi_k \int_{t_{k-1}}^{t_k} (t_i - s)^\alpha \, \mathrm{d}s \right.$$
$$\left. + \int_{t_{i-1}}^{t_i} (t_i - s)^\alpha \, \mathrm{d}W_s - \int_{t_{j-1}}^{t_j} (t_j - s)^\alpha \, \mathrm{d}W_s \right|^2$$

$$\lesssim \left| \sqrt{n} \sum_{k=1}^{j-1} \xi_k \int_{t_{k-1}}^{t_k} \left((t_i - s)^\alpha - (t_j - s)^\alpha\right) \mathrm{d}s \right|^2 + \left| \sqrt{n} \sum_{k=j}^{i-1} \xi_k \int_{t_{k-1}}^{t_k} (t_i - s)^\alpha \, \mathrm{d}s \right|^2$$

$$+ \left| \int_{t_{i-1}}^{t_i} (t_i - s)^\alpha \, \mathrm{d}W_s - \int_{t_{j-1}}^{t_j} (t_j - s)^\alpha \, \mathrm{d}W_s \right|^2.$$

For the first term, since the sequence (ξ_k) is i.i.d. $\sim \mathcal{N}(0, \frac{1}{n})$, we can write

$$\mathbb{E}\left[\left|\sqrt{n}\sum_{k=1}^{j-1}\xi_k\int_{t_{k-1}}^{t_k}\left((t_i-s)^\alpha-(t_j-s)^\alpha\right)ds\right|^2\right]$$

$$=\sum_{k=1}^{j-1}\left|\int_{t_{k-1}}^{t_k}\left((t_i-s)^\alpha-(t_j-s)^\alpha\right)ds\right|^2$$

$$\leq\sum_{k=1}^{j-1}\left|\int_{t_{k-1}}^{t_k}(t_i-t_j)^\alpha ds\right|^2\leq\frac{j-1}{n^2}(t_i-t_j)^{2\alpha}\lesssim(t_i-t_j)^{2\alpha}.$$

For the second term, we have by Jensen's inequality that

$$\mathbb{E}\left[\left|\sqrt{n}\sum_{k=j}^{i-1}\xi_k\int_{t_{k-1}}^{t_k}(t_i-s)^\alpha ds\right|^2\right]=\sum_{k=j}^{i-1}\left|\int_{t_{k-1}}^{t_k}(t_i-s)^\alpha ds\right|^2$$

$$\leq\sum_{k=j}^{i-1}\int_{t_{k-1}}^{t_k}(t_i-s)^{2\alpha}ds$$

$$=\int_{t_{j-1}}^{t_{i-1}}(t_i-s)^{2\alpha}ds$$

$$=\frac{(t_i-t_{j-1})^{2\alpha+1}-(t_i-t_{i-1})^{2\alpha+1}}{2\alpha+1}$$

$$=\frac{1}{2\alpha+1}\left(\left(t_i-t_j+\frac{1}{n}\right)^{2\alpha+1}-\frac{1}{n^{2\alpha+1}}\right)$$

$$\lesssim\frac{(t_i-t_j)^{2\alpha+1}}{2\alpha+1}.$$

Finally, for the last term,

$$
\mathbb{E}\left[\left|\int_{t_{i-1}}^{t_i} (t_i - s)^\alpha dW_s - \int_{t_{j-1}}^{t_j} (t_j - s)^\alpha dW_s\right|^2\right]
$$

$$
= \mathbb{E}\left[\left|\int_{t_{j-1}}^{t_j} \left((t_i - s)^\alpha - (t_j - s)^\alpha\right) dW_s\right.\right.
$$

$$
\left.\left. + \int_{t_j}^{t_i} (t_i - s)^\alpha dW_s - \int_{t_{j-1}}^{t_{i-1}} (t_i - s)^\alpha dW_s\right|^2\right]
$$

$$
\lesssim \mathbb{E}\left[\left|\int_{t_{j-1}}^{t_j} \left((t_i - s)^\alpha - (t_j - s)^\alpha\right) dW_s\right|^2\right]
$$

$$
+ \mathbb{E}\left[\left|\int_{t_j}^{t_i} (t_i - s)^\alpha dW_s\right|^2\right] + \mathbb{E}\left[\left|\int_{t_{j-1}}^{t_{i-1}} (t_i - s)^\alpha dW_s\right|^2\right]
$$

$$
= C\left(\int_{t_{j-1}}^{t_j} \left((t_i - s)^\alpha - (t_j - s)^\alpha\right)^2 ds + \int_{t_j}^{t_i} (t_i - s)^{2\alpha} ds + \int_{t_{j-1}}^{t_{i-1}} (t_i - s)^{2\alpha} ds\right)
$$

$$
\leq C\left(\int_{t_{j-1}}^{t_j} \left((t_i - s)^\alpha - (t_j - s)^\alpha\right)^2 ds + \frac{(t_i - t_j)^{2\alpha+1}}{2\alpha + 1} + \frac{(t_{i-1} - t_{j-1})^{2\alpha+1}}{2\alpha + 1}\right)
$$

$$
\leq C\left(\int_{t_{j-1}}^{t_j} (t_i - t_j)^{2\alpha} ds + 2\frac{(t_i - t_j)^{2\alpha+1}}{2\alpha + 1}\right)
$$

$$
\lesssim (t_i - t_j)^{2\alpha+1}.
$$

Thus by standard moment properties of Gaussian random variables (see Boucheron et al. [14, Theorem 2.1]), we obtain

$$
\mathbb{E}[|\widetilde{\mathcal{G}}^\alpha W_n(t_i) - \widetilde{\mathcal{G}}^\alpha W_n(t_j)|^{2p}] \lesssim \widetilde{\sigma}^{2p} \lesssim (t_i - t_j)^{2p\alpha + p},
$$

which gives the desired result. $\qquad\square$

We further note that Proposition 4.1 and Theorem 2.11 ensure the weak convergence of the log-stock price for the hybrid scheme as well.

Remark 4.2 Proposition 4.1 may easily be extended to a d-dimensional Brownian motion W (for example for multifactor volatility models), also providing a weak convergence result for the d-dimensional version of the hybrid scheme recently developed by Heinrich et al. [37].

## 4.2 Application to fractional binomial trees

We consider a binomial setting for the Riemann–Liouville fractional Brownian motion $\mathcal{G}^{H-1/2} W$ with $g(u) := u^{H-1/2}$, $H \in (0, 1)$, for which Theorem 3.8 provides a weakly converging sequence. On the partition $\mathcal{T}_n$ and with a Bernoulli sequence $(\zeta_i)_{i=1,...,n}$ satisfying $\mathbb{P}[\zeta_i = 1] = \mathbb{P}[\zeta_i = -1] = \frac{1}{2}$ for all i (satisfying the

assumption of Theorem 2.11), the approximating sequence reads

$$(\mathcal{G}^{H-1/2}W_n)(t_i) := \frac{1}{\sqrt{n}}\sum_{k=1}^{i}\zeta_k(t_i - t_{k-1})^{H-1/2} \qquad \text{for } i = 0, \dots, n.$$

Figure 1 shows a fractional binomial tree structure for $H = 0.75$ and $H = 0.1$. Despite being symmetric, such trees cannot be recombining due to the (non-Markovian) path-dependent nature of the process. It might be possible in principle to modify the tree at each step to make it recombining, following the procedure developed in Akyıldırım et al. [1] for Markovian stochastic volatility models. This is not so straightforward, though, and requires a thorough analysis which we leave for future research.

Fig. 1 Binomial tree for the Riemann–Liouville fractional Brownian motion with $n = 5$ discretisation points for $H = 0.75$ (top) and $H = 0.1$ (bottom)
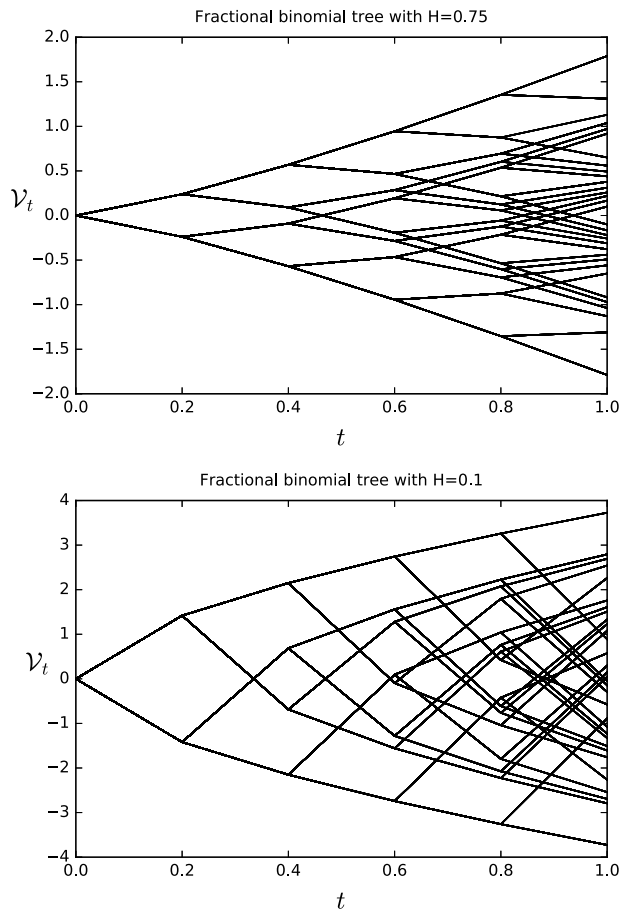
4.3 Monte Carlo

Theorem 2.11 introduces the theoretical foundations for Monte Carlo methods (in particular for path-dependent options) for rough volatility models. In this section, we give a general and easy-to-understand recipe to implement the class of rough volatility models (2.3). For the numerical recipe, we consider the general time partition $\mathcal{T} := \{t_i = \frac{iT}{n} : i = 0, 1, \ldots, n\}$ on $[0, T]$ with $T > 0$.

Algorithm 4.3

1) *Simulate for $j = 1, \ldots, M$ and $i = 1, \ldots, n$ random variables $\xi_{j,i}$, i.i.d., and $\zeta_{j,i}$, i.i.d., all $\sim \mathcal{N}(0, 1)$ and with $\mathrm{corr}(\xi_{j,i}, \zeta_{j,i}) = \rho$.*
2) *Simulate M paths of Y_n via*

$$Y_n^j(t_i) = \frac{T}{n} \sum_{k=1}^{i} b\big(Y_n^j(t_{k-1})\big) + \frac{T}{\sqrt{n}} \sum_{k=1}^{i} a\big(Y_n^j(t_{k-1})\big)\zeta_{j,k}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, M$. Here $Y_n^j(t_i)$ denotes the jth path Y_n evaluated at the time point t_i, which is different from the notation Y_n^j in the theoretical framework above, but should not create any confusion. We also compute, for $i = 1, \ldots, n$ and $j = 1, \ldots, M$,

$$\Delta Y_n^j(t_i) := Y_n^j(t_i) - Y_n^j(t_{i-1}).$$

3) *Simulate M paths of the fractional driving process $((\mathcal{G}^\alpha Y_n)(t))_{t\in\mathcal{T}}$ using*

$$(\mathcal{G}^\alpha Y_n)^j(t_i) := \sum_{k=1}^{i} g(t_{i-k+1})\Delta Y_n^j(t_k) = \sum_{k=1}^{i} g(t_k)\Delta Y_n^j(t_{i-k+1})$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, M$. The complexity of this step is in general of the order $\mathcal{O}(n^2)$ (see Appendix B for details). However, it is easily implemented by using discrete convolutions with complexity $\mathcal{O}(n \log n)$ (see Algorithm B.4 in Appendix B for details in the implementation). With $\mathfrak{g} := (g(t_i))_{i=1,\ldots,n}$ and $\Delta Y_n^j := (\Delta Y_n^j(t_i))_{i=1,\ldots,n}$ for $j = 1, \ldots, M$, we can write $(\mathcal{G}^\alpha Y_n)^j(\mathcal{T}) = \sqrt{\frac{T}{n}}(\mathfrak{g} * \Delta Y_n^j)$ for $j = 1, \ldots, M$, where $*$ represents the discrete convolution operator.

4) *Use the forward Euler scheme to simulate, for $i = 1, \ldots, n$ and $j = 1, \ldots, M$, the log-stock process as*

$$X^j(t_i) = X^j(t_{i-1}) - \frac{1}{2}\frac{T}{n}\Phi(\mathcal{G}^\alpha Y_n)^j(t_{i-1}) + \sqrt{\frac{T}{n}}\sqrt{\Phi(\mathcal{G}^\alpha Y_n)^j(t_{i-1})}\,\xi_{j,i}.$$

Remark 4.4 1) When $Y = W$, we may skip Step 2 in Algorithm 4.3 and replace $\Delta Y_n^j(t_i)$ by $\sqrt{\frac{T}{n}}\zeta_{i,j}$ in Step 3.

2) Step 3 may be replaced by the hybrid scheme from [10] only when $Y = W$.

Antithetic variates in Algorithm 4.3 are easy to implement as it suffices to consider for $j = 1, \ldots, M$ the uncorrelated random vectors $\zeta_j := (\zeta_{j,1}, \zeta_{j,2}, \ldots, \zeta_{j,n})$ and $\xi_j := (\xi_{j,1}, \xi_{j,2}, \ldots, \xi_{j,n})$. Then $(\rho\xi_j + \overline{\rho}\zeta_j, \xi_j), (\rho\xi_j - \overline{\rho}\zeta_j, \xi_j), (-\rho\xi_j - \overline{\rho}\zeta_j, -\xi_j)$ and $(-\rho\xi_j + \overline{\rho}\zeta_j, -\xi_j)$ for $j = 1, \ldots, M$ constitute the antithetic variates, which significantly improves the performance of Algorithm 4.3 by reducing memory requirements, reducing variance and accelerating execution by exploiting symmetry of the antithetic random variables.

4.3.1 Enhancing the performance

A standard practice in Monte Carlo simulation is to match moments of the approximating sequence with the target process. In particular, when the process is Gaussian, matching first and second moments suffices. We only illustrate this approximation for Brownian motion: the left-point approximation (3.4) (with $Y = W$) may be modified to match moments as

$$(\mathcal{G}^\alpha W)(t_i) \approx \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{i} g(t_k^*)\zeta_k \qquad \text{for } i = 0, \ldots, n, \tag{4.1}$$

where t_k^* is chosen optimally. The first two moments of $\mathcal{G}^\alpha W$ read

$$\mathbb{E}[(\mathcal{G}^\alpha W)(t)] = 0, \qquad \text{Var}[(\mathcal{G}^\alpha W)(t)] = \int_0^t g(t-s)^2 ds.$$

The first moment of the approximating sequence (4.1) is always zero, and the second moment reads

$$\text{Var}\left[\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{j-1} g(t_k^*)\zeta_k\right] = \frac{1}{n}\sum_{k=1}^{j-1} g(t_k^*)^2.$$

Equating theoretical and approximating variances gives $\frac{1}{n}g(t_k^*)^2 = \int_{t_{k-1}}^{t_k} g(t-s)^2 ds$ for $k = 1, \ldots, n$, so that the optimal evaluation point can be computed via

$$g(t_k^*) = \sqrt{n \int_{t_{k-1}}^{t_k} g(t-s)^2 ds} \qquad \text{for } k = 1, \ldots, n. \tag{4.2}$$

With the optimal evaluation point, the scheme is still a convolution so that Algorithm B.4 in Appendix B can still be used for faster computations. In the Riemann–Liouville fractional Brownian motion case, $g(u) = u^{H-1/2}$ and the optimal point can be computed in closed form as

$$t_k^* = \left(\frac{n}{2H}\left((t-t_{k-1})^{2H} - (t-t_k)^{2H}\right)\right)^{1/(2H-1)} \qquad \text{for } k = 1, \ldots, n.$$

Proposition 4.5 *The moment matching sequence* $(\widehat{\mathcal{G}}^\alpha W_n)_{n \geq 1}$ *is defined as*

$$\widehat{\mathcal{G}}^\alpha W_n(t) := \widehat{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor}{n}\right)$$

$$+ (nt - \lfloor nt \rfloor)\left(\widehat{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor + 1}{n}\right) - \widehat{\mathcal{G}}^\alpha W_n\left(\frac{\lfloor nt \rfloor}{n}\right)\right) \qquad (4.3)$$

for $t \in \mathbb{I}$, *where*

$$\widehat{\mathcal{G}}^\alpha W_n(t_i) := \sum_{k=1}^{i} \xi_k \sqrt{n \int_{t_{k-1}}^{t_k} g(t_i - s)^2 \mathrm{d}s} \qquad (4.4)$$

with $(\xi_k)_{k \geq 1}$ *an i.i.d. family of centered sub-Gaussian random variables satisfying* $\mathbb{E}[\xi_k^2] = \frac{1}{n}$ *(namely,* $\mathbb{P}[|\xi_k| > x] \leq C\mathrm{e}^{-vx^2}$ *for all* $x > 0$ *and some* $C, v > 0$*). Then weak convergence to* $\mathcal{G}^\alpha W$ *holds in* $(C^{\alpha+1/2}, \| \cdot \|_{\alpha+1/2})$ *for* $\alpha \in (-\frac{1}{2}, \frac{1}{2})$.

Proof Finite-dimensional convergence follows from the central limit theorem as the target process is centered Gaussian; thus convergence of the covariance matrices is enough. It then suffices to prove that the approximating sequence is tight in the desired space, which in view of Theorem 3.4 can be deduced by establishing the control

$$\mathbb{E}[|\widehat{\mathcal{G}}^\alpha W_n(t_i) - \widetilde{\mathcal{G}}^\alpha W_n(t_j)|^{2p}] \leq C|t_i - t_j|^{2p\alpha + p}$$

for all $t_i, t_j \in \mathcal{T}_n$, for $p \geq 1$ and some constant $C \geq 0$. We have

$$\widetilde{\sigma}^2 := \mathbb{E}\left[\left(\widehat{\mathcal{G}}^\alpha W_n(t_i) - \widehat{\mathcal{G}}^\alpha W_n(t_j)\right)^2\right]$$

$$= \mathbb{E}\left[\left(\widehat{\mathcal{G}}^\alpha W_n(t_i)\right)^2\right] + \mathbb{E}\left[\left(\widehat{\mathcal{G}}^\alpha W_n(t_j)\right)^2\right] - 2\mathbb{E}[\widehat{\mathcal{G}}^\alpha W_n(t_i)\widehat{\mathcal{G}}^\alpha W_n(t_j)].$$

We note that

$$\mathbb{E}\left[\left(\widehat{\mathcal{G}}^\alpha W_n(t_i)\right)^2\right] = \sum_{k=1}^{i}\left|\sqrt{\int_{t_{k-1}}^{t_k} g(t_i - s)^2 \mathrm{d}s}\right|^2 = \int_0^{t_i} g(t_i - s)^2 \mathrm{d}s = \mathbb{E}\left[\left(\mathcal{G}^\alpha W(t_i)\right)^2\right],$$

and by Cauchy–Schwarz, we also have

$$\mathbb{E}[\widehat{\mathcal{G}}^\alpha W_n(t_i)\widehat{\mathcal{G}}^\alpha W_n(t_j)] = \sum_{k=1}^{t_i \wedge t_j} \sqrt{\int_{t_{k-1}}^{t_k} g(t_i - s)^2 \mathrm{d}s \int_{t_{k-1}}^{t_k} g(t_j - s)^2 \mathrm{d}s}$$

$$\geq \int_0^{t_i \wedge t_j} g(t_i - s)g(t_j - s)\mathrm{d}s$$

$$= \mathbb{E}[\mathcal{G}^\alpha W(t_i)\mathcal{G}^\alpha W(t_j)].$$

We then obtain $\widetilde{\sigma}^2 \leq \mathbb{E}[|\mathcal{G}^\alpha W(t_i) - \mathcal{G}^\alpha W(t_j)|^2]$. Finally, $\widetilde{\mathcal{G}}^\alpha W_n(t_i) - \widetilde{\mathcal{G}}^\alpha W_n(t_j)$ is sub-Gaussian as a linear combination of sub-Gaussian random variables, and the Gaussian moment inequality (see Boucheron et al. [14, Theorem 2.1]) with the variance estimate $\widetilde{\sigma}^2$ yields

$$\mathbb{E}[|\widehat{\mathcal{G}}^\alpha W_n(t_i) - \widehat{\mathcal{G}}^\alpha W(t_j)|^{2p}] \leq \mathbb{E}[|\mathcal{G}^\alpha W(t_i) - \mathcal{G}^\alpha W(t_j)|^{2p}] \leq \widetilde{C}|t_i - t_j|^{2p\alpha + p}. \qquad \square$$

4.3.2 Reducing variance

As Bayer et al. [6] and Bennedsen et al. [10] pointed out, a major drawback in simulating rough volatility models is the very high variance of the estimators, so that a large number of simulations are needed to produce a decent price estimate. Nevertheless, the rDonsker scheme admits a very simple conditional expectation technique which reduces both memory requirements and variance while also admitting antithetic variates. This approach is best suited for calibrating European type options. We consider $\mathcal{F}_t^B = \sigma(B_s : s \leq t)$ and $\mathcal{F}_t^W = \sigma(W_s : s \leq t)$, the natural filtrations generated by the Brownian motions B and W. In particular, the conditional variance process $(V_t | \mathcal{F}_t^W)$ is deterministic. As discussed by Romano and Touzi [64] and recently adapted to the rBergomi case by McCrickerd and Pakkanen [54], we can decompose the stock price process as

$$e^{X_t} = \mathcal{E}\left(\rho \int_0^t \sqrt{\Phi(\mathcal{G}^\alpha Y)(s)}\, dW_s\right)\mathcal{E}\left(\overline{\rho} \int_0^t \sqrt{\Phi(\mathcal{G}^\alpha Y)(s)}\, dW_s^\perp\right) =: e^{X_t^{||}} e^{X_t^\perp}$$

and notice that

$$X_t | (\mathcal{F}_t^W \vee \mathcal{F}_0^B) \sim \mathcal{N}\left(X_t^{||} - \overline{\rho}^2 \int_0^t \Phi(\mathcal{G}^\alpha Y)(s)ds, \overline{\rho}^2 \int_0^t \Phi(\mathcal{G}^\alpha Y)(s)ds\right).$$

Thus $\exp(X_t)$ becomes lognormal and the Black–Scholes closed-form formulae are valid here (European, barrier options, maximum, etc.). The advantage of this approach is that the orthogonal Brownian motion $W^\perp$ is completely unnecessary for the simulation; hence the generation of random numbers is reduced to a half, yielding proportional memory saving. Not only do we get this, but this simple trick also reduces the variance of the Monte Carlo estimate; hence fewer simulations are needed to obtain the same precision. The following algorithm implements this idea, assuming that $Y = W$.

Algorithm 4.6 *On the equidistant grid $\mathcal{T}$:*

1) Simulate i.i.d. random variables $\zeta_{j,i} \sim \mathcal{N}(0, 1)$ and create antithetic variates $-\zeta_{j,i}$ for $j = 1, \ldots, M/2, i = 1, \ldots, n$.

2) Simulate M paths of the fractional driving process $\mathcal{G}^\alpha W$ using discrete convolution (see Algorithm B.4 in Appendix B for details) to get

$$(\mathcal{G}^\alpha W)^j(\mathcal{T}) = \sqrt{\frac{T}{n}}(\mathfrak{g} * \zeta_j), \qquad j = 1, \ldots, M,$$

and store in memory $\overline{\rho}^2 \int_0^T (\mathcal{G}^\alpha W)^j(s)ds \approx \overline{\rho}^2 \frac{T}{n} \sum_{k=0}^{n-1} (\mathcal{G}^\alpha W)^j(t_k) =: \Sigma^j$ for each $j = 1, \ldots, M$.

3) Use the forward Euler scheme to simulate for $i = 1, \ldots, n, j = 1, \ldots, M$ the log-stock process as

$$X^j(t_i) = X^j(t_{i-1}) - \frac{\rho^2}{2}\frac{T}{n}\Phi(\mathcal{G}^\alpha W)^j(t_{i-1}) + \rho\sqrt{\frac{T}{n}}\sqrt{\Phi(\mathcal{G}^\alpha W)^j(t_{i-1})}\, \zeta_{j,i}.$$

4) *Finally, we may compute any option price using the Black–Scholes formula. For instance, the price of a call option with strike K and maturity $T \in \mathbb{I}$ would be given by $C^j(K) = \exp(X^j(T))\mathcal{N}(d_1^j) - K\mathcal{N}(d_2^j)$ for $j = 1, \dots, M$, with $d_{1,2}^j$ given by $d_1^j := \frac{1}{\sqrt{\Sigma^j}}(X^j(T) - \log K + \frac{1}{2}\Sigma^j)$ and $d_2^j = d_1^j - \sqrt{\Sigma^j}$. Thus the output of the model would be $C(K) = \frac{1}{M}\sum_{k=1}^M C^j(K)$.*

The algorithm is easily adapted to general diffusions Y as drivers of the volatility (Algorithm 4.3, Step 2). Algorithm 4.3 is obviously faster than Algorithm 4.6, especially when using control variates. Nevertheless, with the same number of paths, Algorithm 4.6 remarkably reduces the Monte Carlo variance, meaning in turn that fewer simulations are needed, making it very competitive for calibration.

4.4 Numerical example: rough Bergomi model

Figures 2–5 perform a numerical analysis of the Monte Carlo convergence as a function of n. We observe that the lower H, the larger n needs to be to achieve convergence. However, we also observe that for the Cholesky, rDonsker (naive and moment match) and hybrid schemes and $H \geq 0.1$, with $n = 252$, we already achieve a precision of the order 10^{-4}, which is equivalent to a basis point in financial terms. For $H < 0.1$, we might require n larger than 252 if precision is required beyond 10^{-4}. We also observe in Fig. 5 that the naive rDonsker approximation converges extremely slowly for small H. Additionally, Figs. 6–11 measure the price estimations compared to the Cholesky method which is taken as benchmark. The hybrid scheme tends to
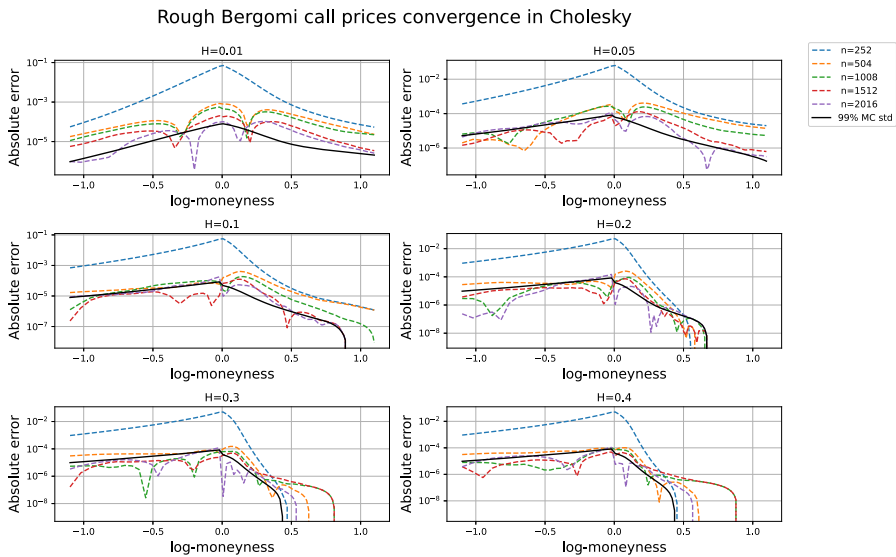
Fig. 2 Rough Bergomi call option price convergence using the Cholesky method with $\xi_0 = 0.04$, $\nu = 2.3$, $\rho = -0.9$, $S_0 = 1$, $T = 1$ with 2×10^6 simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$, the previous discretisation is $n = 126$
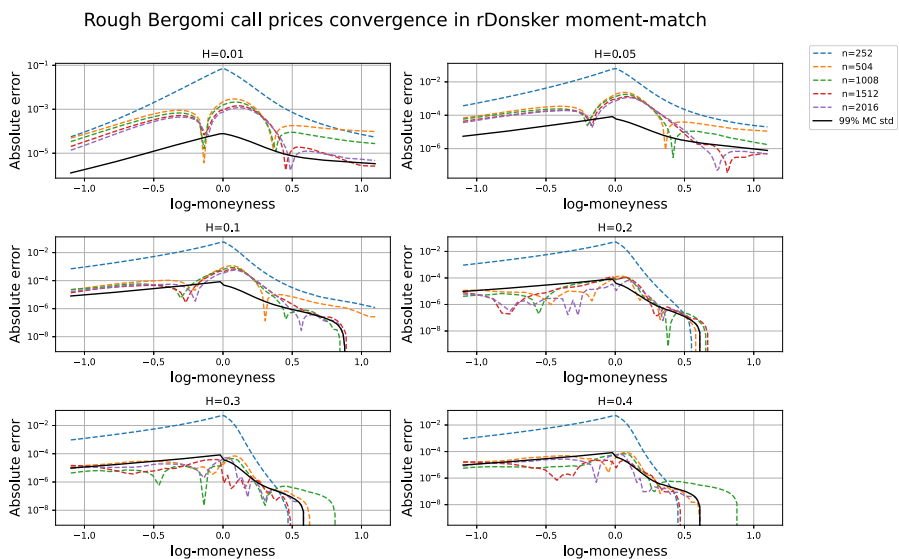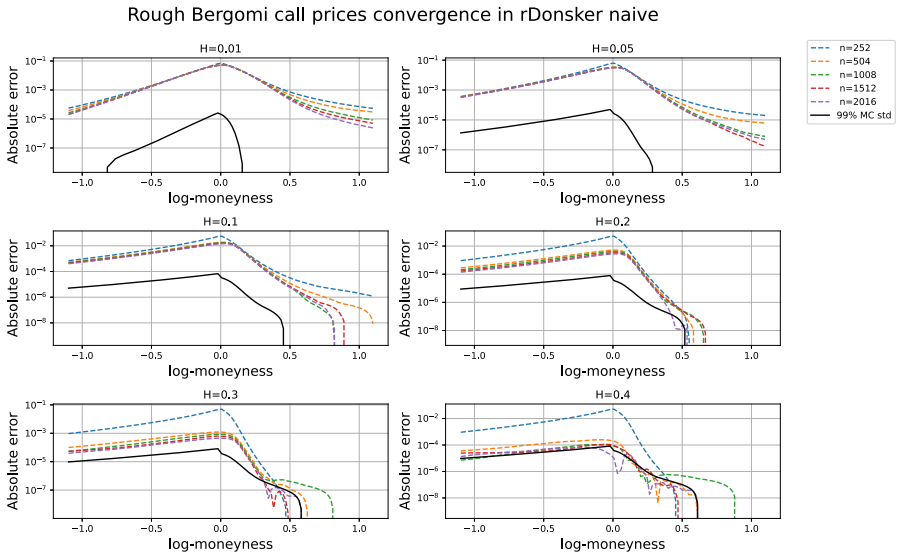
Rough Bergomi call prices convergence in Hybrid
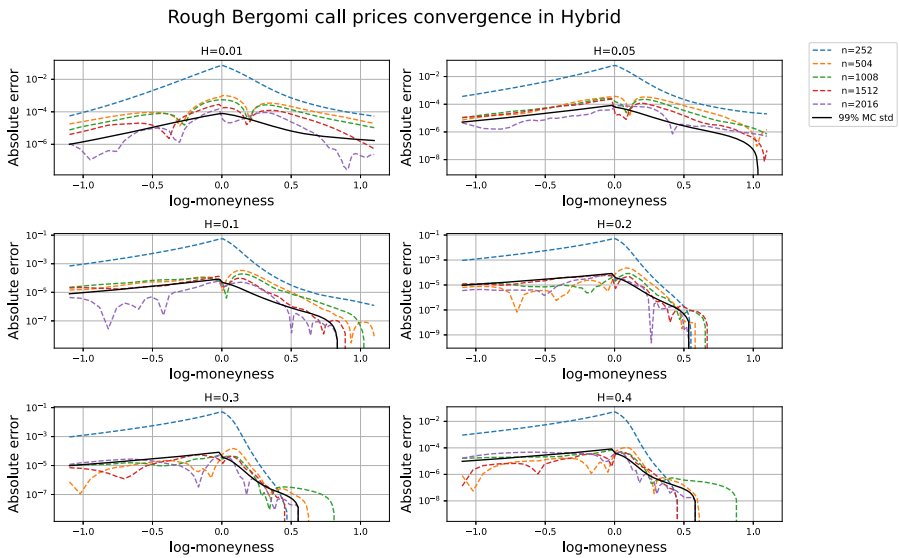
Fig. 3 Rough Bergomi call option price convergence using the hybrid scheme with $\xi_0 = 0.04$, $\nu = 2.3$, $\rho = -0.9$, $S_0 = 1$, $T = 1$ with 2×10^6 simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$, the previous discretisation is $n = 126$

Rough Bergomi call prices convergence in rDonsker moment-match

Fig. 4 Rough Bergomi call option price convergence using rDonsker with moment matching and $\xi_0 = 0.04$, $\nu = 2.3$, $\rho = -0.9$, $S_0 = 1$, $T = 1$ with 2×10^6 simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$, the previous discretisation is $n = 126$

Rough Bergomi call prices convergence in rDonsker naive

Fig. 5 Rough Bergomi call option price convergence using rDonsker method with left-point Euler and $\xi_0 = 0.04$, $\nu = 2.3$, $\rho = -0.9$, $S_0 = 1$, $T = 1$ with 2×10^6 simulations and antithetic variates. Absolute error represents the difference between subsequent approximations, where n represents the time grid size. For $n = 252$, the previous discretisation is $n = 126$

Price estimate error for H=0.01 with respect to Cholesky
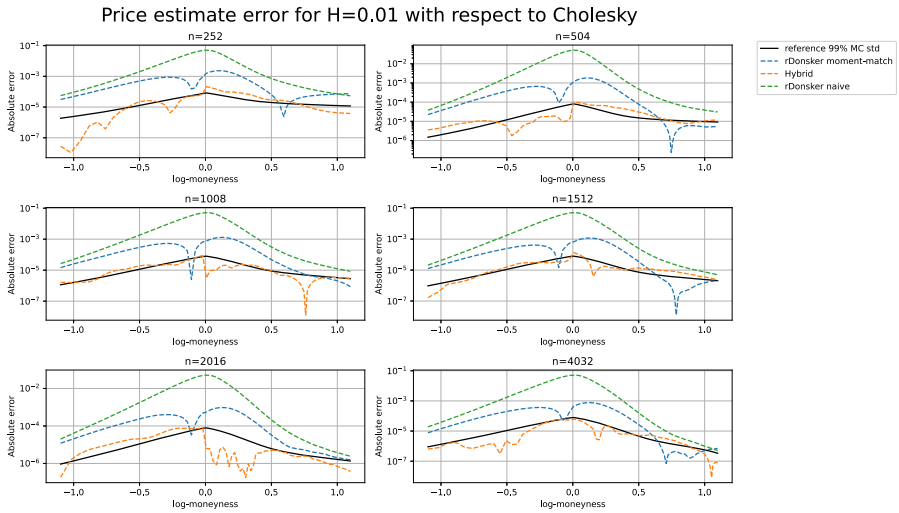
Fig. 6 Rough Bergomi call option price comparison with $H = 0.01$, $\xi_0 = 0.04$, $\nu = 2.3$, $\rho = -0.9$, $S_0 = 1$, $T = 1$ with 2×10^6 simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes
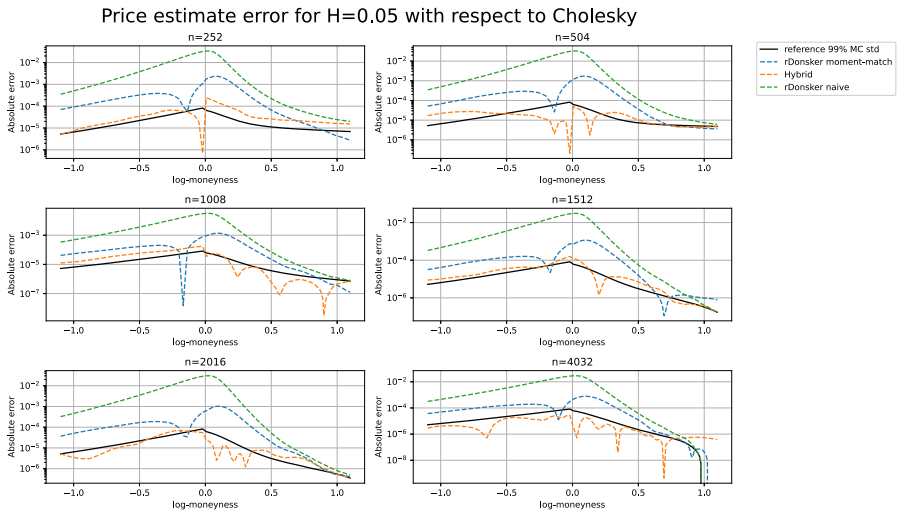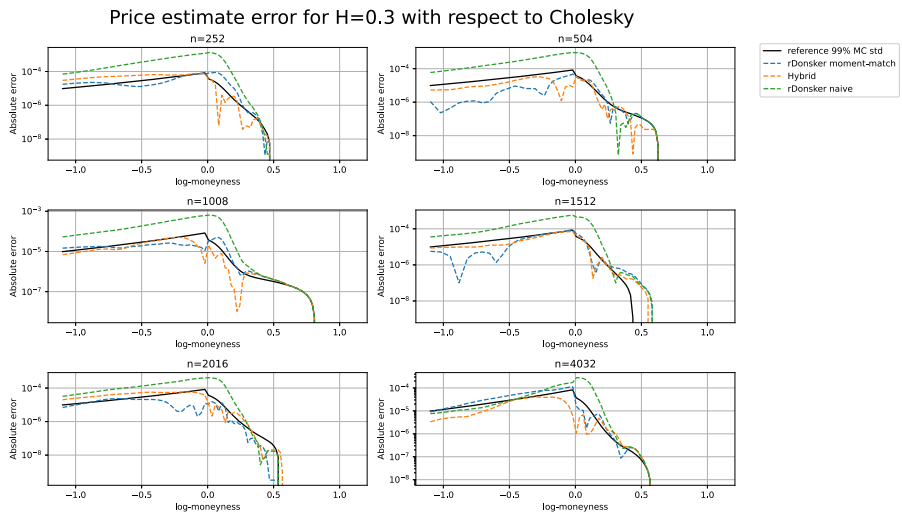
Fig. 7 Rough Bergomi call option price comparison with $H = 0.05$, $\xi_0 = 0.04$, $\nu = 2.3$, $\rho = -0.9$, $S_0 = 1$, $T = 1$ with 2×10^6 simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes
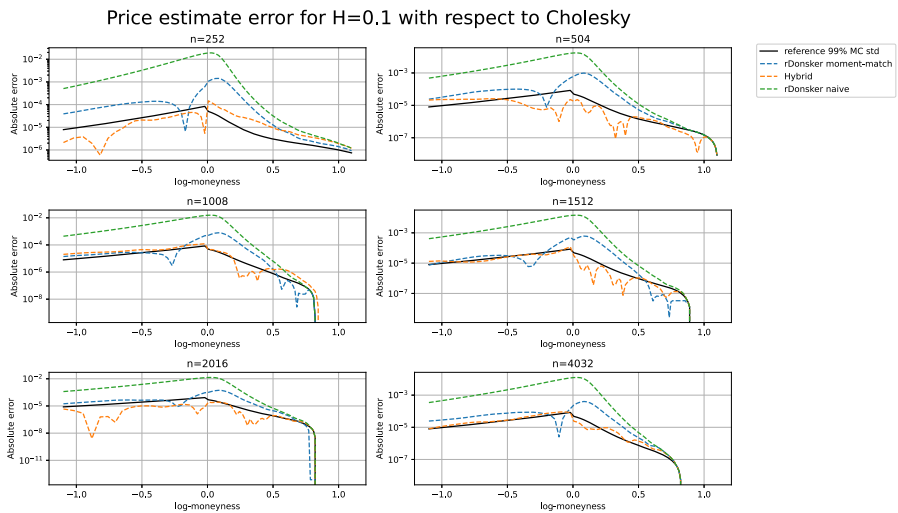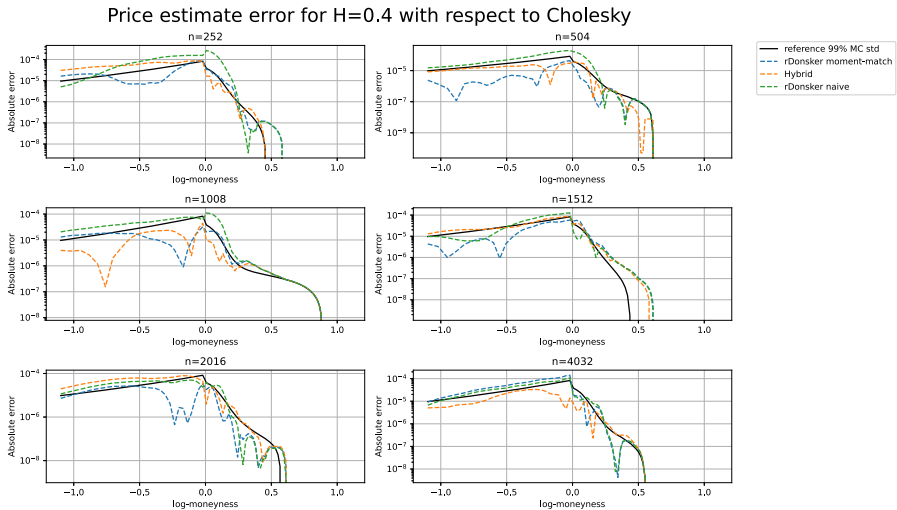

Fig. 8 Rough Bergomi call option price comparison with $H = 0.10$, $\xi_0 = 0.04$, $\nu = 2.3$, $\rho = -0.9$, $S_0 = 1$, $T = 1$ with 2×10^6 simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes

Price estimate error for H=0.2 with respect to Cholesky
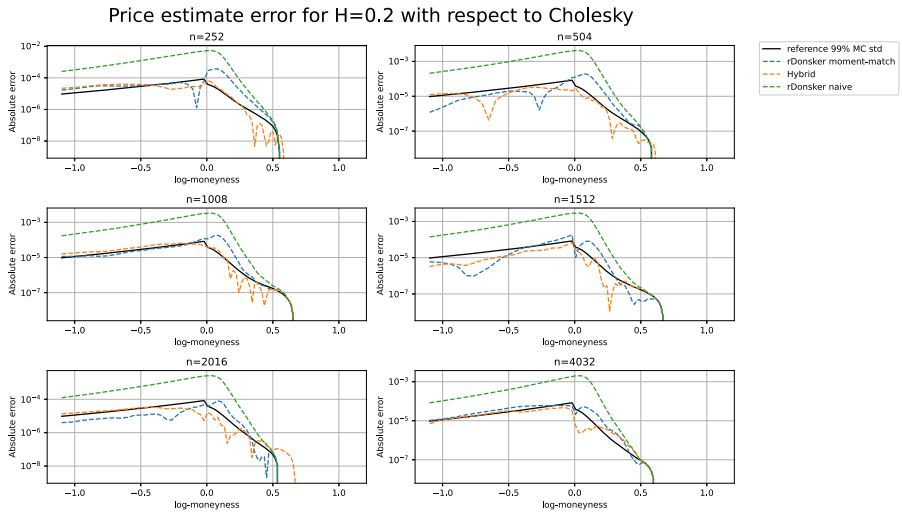
Fig. 9 Rough Bergomi call option price comparison with $H = 0.2$, $\xi_0 = 0.04$, $\nu = 2.3$, $\rho = -0.9$, $S_0 = 1$, $T = 1$ with 2×10^6 simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes

Price estimate error for H=0.3 with respect to Cholesky

Fig. 10 Rough Bergomi call option price comparison with $H = 0.3$, $\xi_0 = 0.04$, $\nu = 2.3$, $\rho = -0.9$, $S_0 = 1$, $T = 1$ with 2×10^6 simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes

Fig. 11 Rough Bergomi call option price comparison with $H = 0.4$, $\xi_0 = 0.04$, $\nu = 2.3$, $\rho = -0.9$, $S_0 = 1$, $T = 1$ with 2×10^6 simulations and antithetic variates. Absolute error represents the difference in price between different simulation schemes

be closer to this benchmark especially for $H < 0.1$. When $H \geq 0.1$, for both the hybrid scheme and rDonsker moment match, we observe an error less than 10^{-4} for $n \geq 252$. It is noteworthy to mention that the naive rDonsker scheme has substantially worse convergence (at least an order of magnitude) than the other methods. We note that the black lines in all figures represent the 99% Monte Carlo standard deviations; hence errors below that threshold should be interpreted as noise.

4.5 Speed benchmark against Markovian stochastic volatility models

In this section, we benchmark the speed of the rDonsker scheme against the hybrid scheme and a classical Markovian stochastic volatility model using 10^5 simulations and averaging the speeds over 10 trials. For the former, we simulate the rBergomi model from Bayer et al. [6], whereas for the latter, we use the classical Bergomi [12] model using a forward Euler scheme in both volatility and stock price. All three schemes are implemented in `Cython` to make the comparison fair and to obtain speeds comparable to `C++`. Figure 12 shows that rDonsker is about 2 times slower than the Markovian case, whereas the hybrid scheme is approximately 2.5 times slower, which is expected from the complexities of the two schemes. However, it is remarkable that the $\mathcal{O}(n \log n)$ complexity of the FFT stays almost constant with the grid size n, and the computational time grows almost linearly as in the Markovian case. We presume that this is the case since $n \ll 10'000$ is relatively small. Figure 12 also shows that rough volatility models can be implemented very efficiently and are not much slower than classical stochastic volatility models.
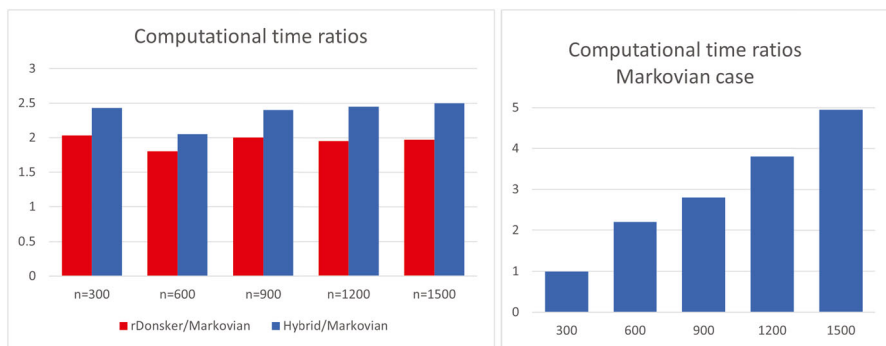
Fig. 12 Computational time benchmark using hybrid scheme, rDonsker and Markovian (forward Euler) for different grid sizes n. The left-hand figure shows the ratios 'hybrid scheme / Markovian case' and 'rDonsker scheme / Markovian case', where the Markovian case refers to a classical Euler scheme in a non-rough setting (i.e., with $H = \frac{1}{2}$). The right-hand figure shows the computational time of the Markovian case only, normalised at 1 for $n = 300$

Table 1 Recommendations for the use of different numerical schemes depending on the Hurst parameter

$H > 0.1$	$H \in [0.05, 0.1]$	$H < 0.05$
rDonsker	choice depends on error sensitivity	hybrid scheme

4.6 Implementation guidelines and conclusion

The numerical analysis above suggests some guidelines to implement rough volatility models driven by TBSS processes of the form $\mathcal{G}^{H-1/2}Y$ for some Itô diffusion Y.

Regarding empirical estimates, Gatheral et al. [32] suggest that $H \approx 0.15$. Bennedsen et al. [11] give an exhaustive analysis of more than 2000 equities for which $H \in [0.05, 0.2]$. On the pricing side, Bayer et al. [6] and Jacquier et al. [40] found that calibration yields $H \in [0.05, 0.10]$. Finally, Livieri et al. [50] found evidence in options data that $H \approx 0.3$. Despite the diverse ranges found so far, there is a common agreement that $H < \frac{1}{2}$. Table 1 provides recommendations about when to use which numerical scheme depending on the value of the Hurst parameter.

Remark 4.7 The rough Heston model by Guennoun et al. [33] is out of the scope of the hybrid scheme. Moreover, any process of the form $\mathcal{G}^{\alpha}Y$ for some Itô diffusion Y under Assumption 2.3 is in general out of the scope of the hybrid scheme. This only leaves the choice of using the rDonsker scheme, for which reasonable accuracy is obtained at least for Hölder regularities greater than 0.05.

4.7 Bushy trees and binomial markets

Binomial trees have attracted a lot of attention from both academics and practitioners, as their apparent simplicity provides easy intuition about the dynamics of a given asset. Not only this, but they are by construction arbitrage-free and allow finding prices of path-dependent options, together with their hedging strategy. In particular, early

exercise options, in particular Bermudan or American options, are usually priced using trees as opposed to Monte Carlo methods. The convergence stated in Theorem 2.11 lays the theoretical foundations to construct fractional binomial trees (note that Bernoulli random variables satisfy the conditions of the theorem). Figure 1 already showed binomial trees for fractional Brownian motion, but we ultimately need trees describing the dynamics of the stock price.

4.7.1 A binary market

We invoke Theorem 2.11 with the independent sequences $(\zeta_i)_{i=1,\dots,n}$, $(\zeta_i^\perp)_{i=1,\dots,n}$ such that $\mathbb{P}[\zeta_i = 1] = \mathbb{P}[\zeta_i^\perp = 1] = \mathbb{P}[\zeta_i = 1] = \mathbb{P}[\zeta_i^\perp = -1] = \frac{1}{2}$ for all i. We further define on $\mathcal{T}$ for any $i = 1,\dots,n$ the quantities

$$B_n(t_i) = \sqrt{\frac{T}{n}} \sum_{k=1}^{i} (\rho \zeta_k + \overline{\rho} \zeta_k^\perp),$$

$$Y_n(t_i) = \frac{T}{n} \sum_{k=1}^{i} b\big(Y_n(t_{k-1})\big) + \sqrt{\frac{T}{n}} \sum_{k=1}^{i} \sigma\big(Y_n(t_{k-1})\big) \zeta_k,$$

the approximating sequences to B and Y in (2.3). The approximation for X is then given by

$$X_n(t_i) = X_n(t_{i-1}) - \frac{1}{2}\frac{T}{n} \sum_{k=1}^{i} \Phi(\mathcal{G}^\alpha Y_n)(t_k) + \sqrt{\frac{T}{n}} \sum_{k=1}^{i} \sqrt{\Phi(\mathcal{G}^\alpha Y_n)(t_k)}\, (\rho \zeta_k + \overline{\rho} \zeta_k^\perp).$$

In order to construct the tree, we have to consider all possible permutations of the random vectors $(\zeta_i)_{i=1,\dots,n}$ and $(\zeta_i^\perp)_{i=1,\dots,n}$. Since each random variable only takes two values, this adds up to 4^n possible combinations; hence the 'bushy tree' terminology. When $\rho \in \{-1, 1\}$, the magnitude is reduced to 2^n.

4.8 American options in rough volatility models

There is so far no available scheme for American options (or any early-exercise options for that matter) under rough volatility models, but the fractional trees constructed above provide a framework to do so. In the Black–Scholes model, American options can be priced using binomial trees by backward induction. A key ingredient is the Snell envelope and the following representation.

Definition 4.8 Let $(X(t))_{t\in\mathbb{I}}$ be an $(\mathcal{F}_t)_{t\in\mathbb{I}}$-adapted process. The Snell envelope $\mathcal{J}$ of X with respect to $(\mathcal{F}_t)_{t\in\mathbb{I}}$ is defined as $\mathcal{J}(X)(t) := \operatorname{esssup}_{\tau\in\tilde{\mathbb{I}}_t} \mathbb{E}[X(\tau)|\mathcal{F}_t]$ for all $t \in \mathbb{I}$, where $\tilde{\mathbb{I}}_t$ denotes the set of stopping times with values in $[t, 1] \subseteq \mathbb{I}$.

Definition 4.9 For a log-strike k and price process S, write $C(t) := \mathrm{e}^{-rt}(S(t) - \mathrm{e}^k)^+$ and $P(t) := \mathrm{e}^{-rt}(\mathrm{e}^k - S(t))^+$. Then the discounted time-t prices of an American call and put with log-strike k and maturity T are given by, respectively,

$$\mathrm{e}^{-rt} C_t^a(k, T) = \mathcal{J}(C)(t), \qquad \mathrm{e}^{-rt} P_t^a(k, T) = \mathcal{J}(P)(t).$$

Preservation of weak convergence under the Snell envelope map is due to Mulinacci and Pratelli [55], who proved that convergence takes place in the Skorokhod topology provided that the Snell envelope is continuous. In our setting, the scheme for American options is justified by the following result. Recall that a filtration is said to be *immersed* in a larger filtration if all local martingales in the smaller filtration are also local martingales in the larger filtration.

Theorem 4.10 *Let V and X in (2.3) be such that $\mathbb{E}[\sup_{t \in \mathbb{I}} e^{X_t}] < \infty$ and the natural filtration of X is immersed in the natural filtration of (B, W) from (2.3) and (2.4). With the sequence $(X_n)_{n \geq 1}$ converging in $(\mathcal{D}(\mathbb{I}), d_{\mathcal{D}})$ weakly to X, as per Theorem 2.11, assume additionally that $(X_n)_{n \geq 1}$ satisfies the conditions of Mulinacci and Pratelli [55, Theorem 3.5]. Let $S = e^X$ and $S_n = e^{X_n}$. If for each $n \geq 1$, C_n and P_n are defined as in Definition 4.9 for the price process S_n, then $(S_n, \mathcal{J}(C_n), \mathcal{J}(P_n))$ converges in $(\mathcal{D}(\mathbb{I}; \mathbb{R}^3), d_{\mathcal{D}})$ weakly to $(S, \mathcal{J}(C), \mathcal{J}(P))$, where the Snell envelopes are defined with respect to the natural filtrations of the corresponding price processes.*

Proof By the continuous mapping theorem, the weak convergence of (X_n) implies that (S_n, C_n, P_n) converges in $(\mathcal{D}(\mathbb{I}; \mathbb{R}^3), d_{\mathcal{D}})$ weakly to (S, C, P). Because C and P are continuous and $\mathbb{E}[\sup_{t \in \mathbb{I}} e^{X_t}] < \infty$, it follows from Karatzas and Shreve [43, Theorem D.13] that for both $\mathcal{J}(C)$ and $\mathcal{J}(P)$, the predictable finite-variation part of the Doob–Meyer decomposition is continuous. Moreover, the assumption that the natural filtration of X is immersed in that of (W, B) implies that also the martingale part is continuous. Therefore $\mathcal{J}(C)$ and $\mathcal{J}(P)$ are continuous processes. In view of the additional assumptions, the claim now follows from [55, Theorem 5.5]. □

Mulinacci and Pratelli [55] also gave explicit conditions for weak convergence to be preserved in the Markovian case. It is trivial to see that the pricing of American options in the rough tree scheme coincides with the classical backward induction procedure. We consider a continuously compounded interest rate r and dividend yield d.

Algorithm 4.11 *On the equidistant grid $\mathcal{T}$:*

1) *Construct the binomial tree using the explicit construction in Sect. 4.7.1 and obtain $(S_t^j)_{t \in \mathcal{T}, j=1,\ldots,4^n}$.*

2) *The backward recursion for the American option price with payoff function $h(\cdot)$ is given by $\widetilde{h}_{t_N} := h(S_{t_N})$ and*

$$\widetilde{h}_{t_i} := e^{(d-r)/n} \mathbb{E}[\widetilde{h}_{t_{i+1}} | \mathcal{F}_{t_i}] \vee h(S_{t_i}) \qquad for\ i = N - 1, \ldots, 0,$$

where $\mathbb{E}[\cdot | \mathcal{F}_{t_i}] = \frac{1}{4}(\widetilde{h}_{t_{i+1}}^{++} + \widetilde{h}_{t_{i+1}}^{+-} + \widetilde{h}_{t_{i+1}}^{-+} + \widetilde{h}_{t_{i+1}}^{--})$ and $\widetilde{h}_{t_i}^{\pm\pm}$ represents the outcome $(\zeta_i, \zeta_i^{\perp}) = (\pm 1, \pm 1)$ for the driving binomials, following the construction in Sect. 4.7.1.

3) *Finally, $\widetilde{h}_0$ is the price of the American option at the inception of the contract.*

The main computational cost of the scheme is the construction of the tree in Step 1). Once the tree is constructed, computing American prices for different options is a fast routine.

4.8.1 Numerical example: rough Bergomi model

We note that, as proved by Gassiat [31, Theorem 1.1(1)], the price process in the rough Bergomi model is a true martingale when $\rho \leq 0$. We construct a rough volatil-
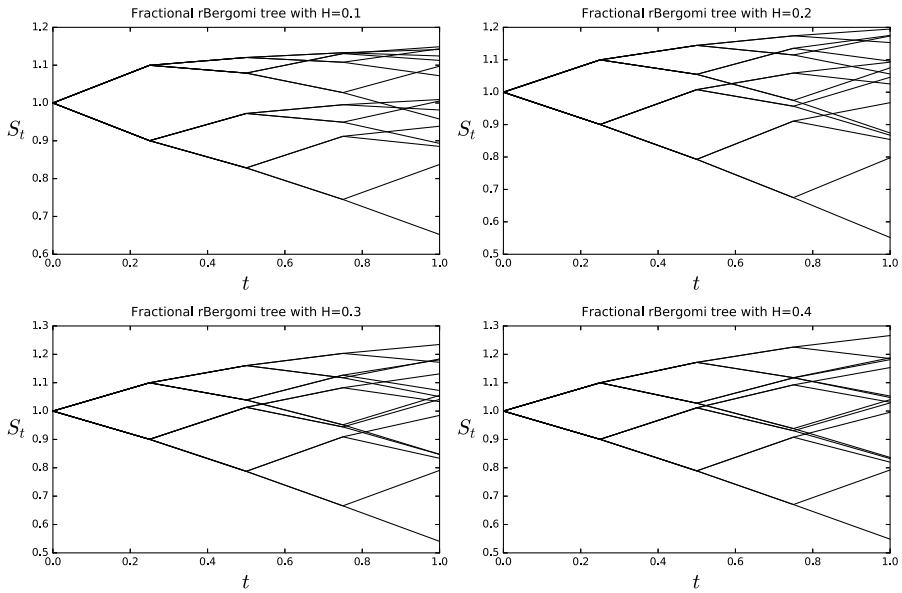

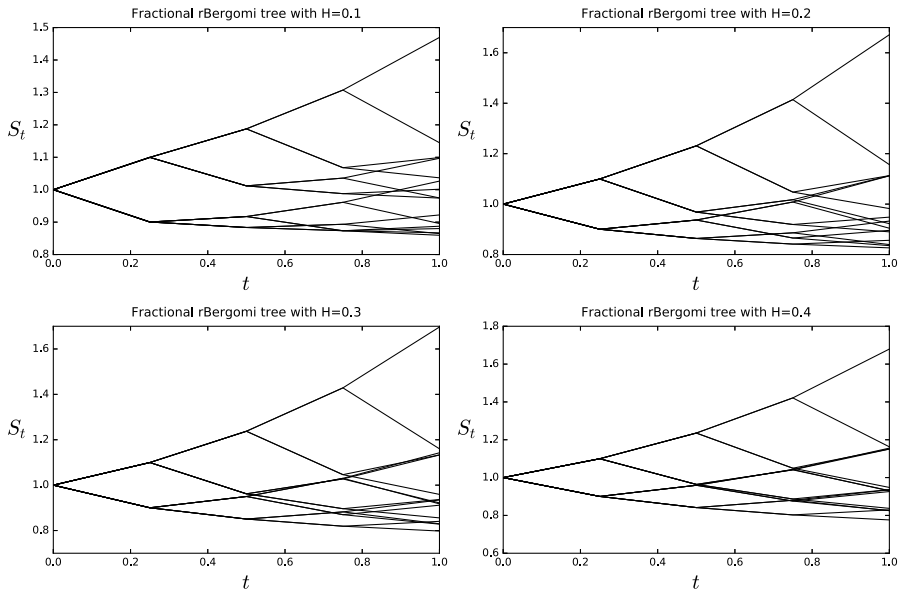**Fig. 13** rBergomi trees for different values of H, with $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ and 5 time steps

Fig. 14 rBergomi trees for different values of H, with $(\nu, \rho, \xi_0) = (1, 1, 0.04)$ and 5 time steps

ity tree for the rough Bergomi model (introduced in [6]) and check the accuracy of the scheme. Figures 13 and 14 show the fractional trees for different values of H and for $\rho \in \{-1, 1\}$. Both pictures show a markedly different behaviour, but as a common property, we observe that as H tends to $1/2$, the tree structure somehow becomes simpler.

4.8.2 European options

Figure 15 displays volatility smiles obtained using the tree scheme. Even though the time steps are not sufficient for small H, the fit remarkably improves when $H \geq 0.15$, and always remains inside the 95% confidence interval with respect to the hybrid scheme. Moreover, the moment-matching approach from Sect. 4.3.1 shows a better accuracy than the standard rDonsker scheme when $H \leq 0.1$. In Fig. 16, a detailed error analysis corroborates these observations: the relative error is smaller than 3% for $H \geq 0.15$.
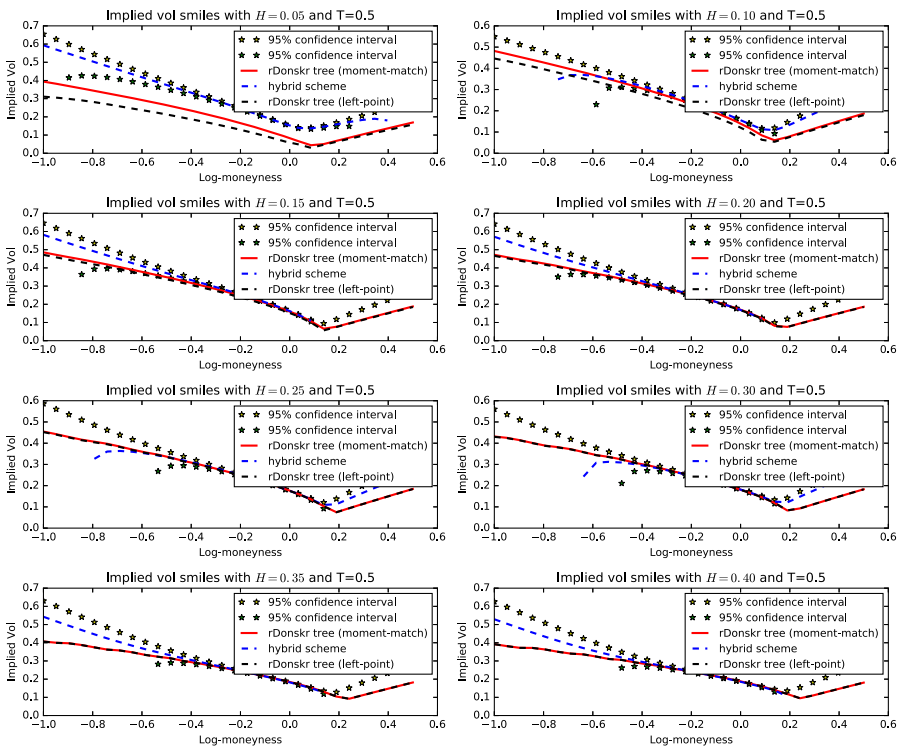
Fig. 15 Implied volatility smiles with varying H, $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ and 24 time steps
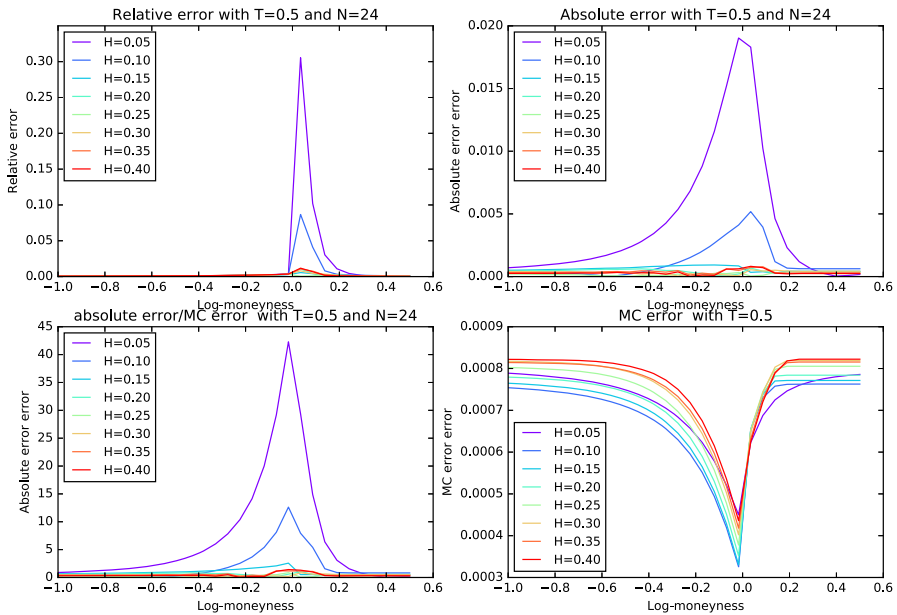
Fig. 16 Error analysis for the rDonsker moment-match tree for different values of H, with $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ and 24 time steps

4.8.3 American options

In the context of American options, there is no benchmark to compare our result. However, the accurate results found in the previous section (at least for $H \geq 0.15$) justify the use of trees to price American options. Figure 17 shows the output of American and European put prices with interest rates equal to $r = 5\%$. Interestingly, the rougher the process (the smaller H), the larger the difference between in-the-money European and American options.
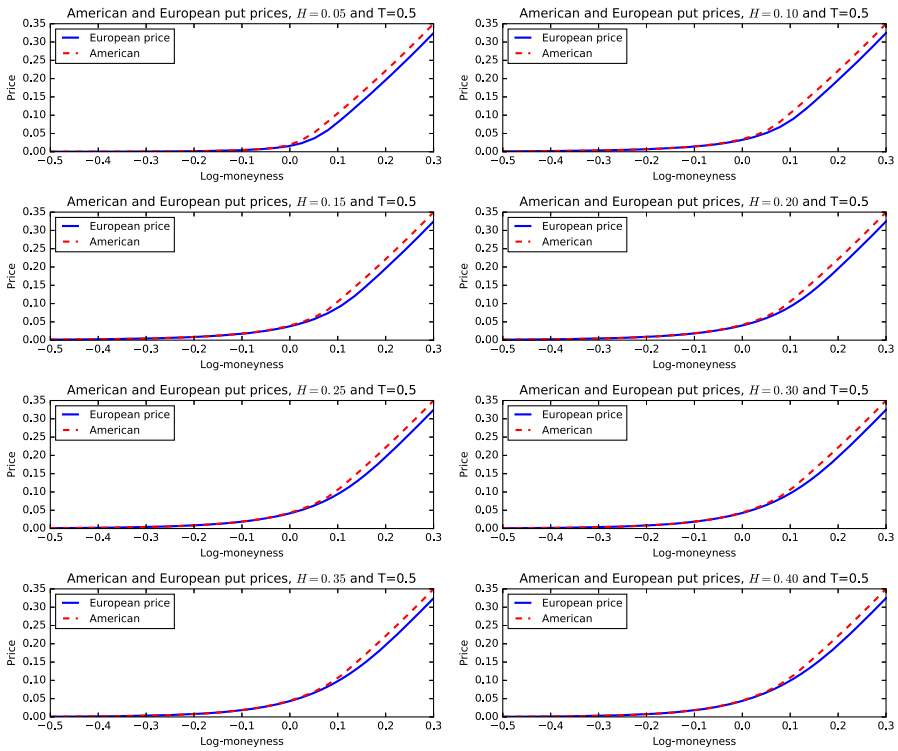
Fig. 17 American and European put prices in the rough Bergomi model for different values of H and $(\nu, \rho, \xi_0) = (1, -1, 0.04)$ with 26 time steps

Appendix A: Riemann–Liouville operators

We review here fractional operators and their mapping properties. We follow closely the excellent monograph by Samko et al. [65, Chap. 1], as well as some classical results by Hardy and Littlewood [36]. However, we introduce a modification in their definition, so that the condition $f(0) = 0$ is not necessary as opposed to the original definition in [36].

A.1 Riemann–Liouville fractional operators

Definition A.1 For $\lambda \in (0, 1)$, $\alpha \in (-\lambda, 1 - \lambda)$, the *left Riemann–Liouville fractional operator* is defined on $\mathcal{C}^\lambda(\mathbb{I})$ as

$$(I^\alpha f)(t) := \begin{cases} \frac{1}{\Gamma(\alpha)} \int_0^t \frac{f(s) - f(0)}{(t-s)^{1-\alpha}} \mathrm{d}s & \text{for } \alpha \in (0, 1 - \lambda), \\ (\frac{\mathrm{d}}{\mathrm{d}t} I^{1+\alpha} f)(t) & \\ \quad = \frac{1}{\Gamma(1+\alpha)} \frac{\mathrm{d}}{\mathrm{d}t} (\int_0^t (t-s)^\alpha (f(s) - f(0)) \mathrm{d}s) & \text{for } \alpha \in (-\lambda, 0). \end{cases} \tag{A.1}$$

Theorem A.2 *For any $f \in \mathcal{C}^\lambda(\mathbb{I})$ with $\lambda \in (0, 1)$ and $\alpha \in (-\lambda, 1 - \lambda)$, we have $I^\alpha f \in \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$. In particular, there exists $C > 0$ such that $|(I^\alpha f)(t)| \leq Ct^{\alpha+\lambda}$ for any $t \in \mathbb{I}$.*

Proof We first consider $\alpha > 0$. Then using the definition and $f \in \mathcal{C}^\lambda(\mathbb{I})$, we obtain

$$|(I^\alpha f)(t)| = \frac{1}{\Gamma(\alpha)} \left| \int_0^t \frac{f(u) - f(0)}{(t-u)^{1-\alpha}} du \right|$$

$$\leq \frac{|f|_\lambda}{\Gamma(\alpha)} \int_0^t \frac{u^\lambda du}{(t-u)^{1-\alpha}}$$

$$\leq \frac{\Gamma(2+\lambda)|f|_\lambda}{(1+\lambda)\Gamma(\alpha+\lambda+1)} t^{\alpha+\lambda},$$

which proves the estimate for $|I^\alpha f|$. Next we prove that $I^\alpha f \in \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$. For this, introduce $\phi(t) := f(t) - f(0)$ and consider $t, t+h \in \mathbb{I}$ with $h > 0$. Then

$$(I^\alpha f)(t+h) - (I^\alpha f)(t)$$

$$= \frac{1}{\Gamma(\alpha)} \left(\int_{-h}^t \frac{\phi(t-u)}{(u+h)^{1-\alpha}} du - \int_0^t \frac{\phi(t-u)}{u^{1-\alpha}} du \right)$$

$$= \frac{\phi(t)}{\Gamma(1+\alpha)} \left((t+h)^\alpha - t^\alpha \right) + \frac{1}{\Gamma(\alpha)} \int_{-h}^0 \frac{\phi(t-u) - \phi(t)}{(u+h)^{1-\alpha}} du$$

$$+ \frac{1}{\Gamma(\alpha)} \int_0^t \left((u+h)^{\alpha-1} - u^{\alpha-1} \right) \left(\phi(t-u) - \phi(t) \right) du$$

$$=: J_1 + J_2 + J_3. \tag{A.2}$$

We first consider J_1. If $h > t$, then

$$|J_1| \leq \frac{|f|_\lambda}{\Gamma(1+\alpha)} t^\lambda \left((t+h)^\alpha - t^\alpha \right) \leq Ch^{\lambda+\alpha}.$$

On the other hand, when $0 < h < t$, since $(1+u)^\alpha - 1 \leq \alpha u$ for $u > 0$, we get

$$|J_1| \leq \frac{|f|_\lambda}{\Gamma(1+\alpha)} t^{\lambda+\alpha} \left| \left(1 + \frac{h}{t} \right)^\alpha - 1 \right| \leq Cht^{\lambda+\alpha-1} \leq Ch^{\lambda+\alpha}.$$

For J_2, since $f \in \mathcal{C}^\lambda(\mathbb{I})$, we can write

$$|J_2| \leq \frac{|f|_\lambda}{\Gamma(\alpha)} \int_{-h}^0 \frac{|u|^\lambda}{(u+h)^{1-\alpha}} \leq Ch^{\lambda+\alpha}.$$

Finally,

$$|J_3| \leq \frac{|f|_\lambda}{\Gamma(\alpha)} \int_0^t u^\lambda \left(u^{\alpha-1} - (u+h)^{\alpha-1} \right) du$$

$$= \frac{|f|_\lambda}{\Gamma(\alpha)} h^{\lambda+\alpha} \int_0^{t/h} u^\lambda \left(u^{\alpha-1} - (u+1)^{\alpha-1} \right) du.$$

Hence if $t \leq h$, then $|J_3| \leq Ch^{\lambda+\alpha}$. Likewise, if $t > h$ and $\lambda + \alpha < 1$, then $|J_3| \leq Ch^{\lambda+\alpha}$ since

$$|u^{\alpha-1} - (u+1)^{\alpha-1}| = u^{\alpha-1}\left(1 - \left(1 + \frac{1}{u}\right)^{\alpha-1}\right) \leq Cu^{\alpha-2}.$$

Thus we have shown that $I^\alpha f$ satisfies the $(\lambda + \alpha)$-Hölder condition and belongs to $\mathcal{C}^{\lambda+\alpha}(\mathbb{I})$ in the case where $\alpha > 0$. The conclusion for $\alpha < 0$ follows by taking $g(u) := u^\alpha$ in the proof of Proposition 2.2 in Appendix C. $\qquad\square$

Corollary A.3 *For any $\lambda \in (0,1)$ and $\alpha \in (-\lambda, 1 - \lambda)$, I^α is a continuous operator from $\mathcal{C}^\lambda(\mathbb{I})$ to $\mathcal{C}^{\lambda+\alpha}(\mathbb{I})$.*

Proof Clearly, I^α is a linear operator. From Theorem A.2, $\|I^\alpha f\|_{\alpha+\lambda} \leq C\|f\|_\lambda$ since $|f|_\lambda \leq \|f\|_\lambda$. Therefore I^α is also bounded and hence continuous. $\qquad\square$

Appendix B: Discrete convolution

Definition B.1 For a, b $\in \mathbb{R}^n$, the *discrete convolution operator* $* : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is defined as

$$(a * b)_i := \sum_{m=0}^{i} a_m b_{i-m}, \qquad i = 0, \dots, n-1.$$

When simulating $\mathcal{G}^\alpha W$ on the uniform partition $\mathcal{T}$, the scheme (Algorithm 4.3, Step 3) reads

$$(\mathcal{G}^\alpha W)^j(t_i) = \sum_{k=1}^{i} g(t_k)\zeta_{j,k-i+1} \qquad \text{for } i = 1, \dots, n,$$

which has the form of the discrete convolution in Definition B.1. Rewritten in matrix form,

$$\begin{pmatrix} g(t_1) & 0 & \dots & 0 \\ g(t_2) & g(t_1) & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ g(t_n) & g(t_{n-1}) & \dots & g(t_1) \end{pmatrix} \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_n \end{pmatrix},$$

it is clear that this operation yields a complexity of the order $\mathcal{O}(n^2)$, which can be improved drastically.

Definition B.2 For a sequence c $:= (c_0, c_1, \dots, c_{n-1}) \in \mathbb{C}^n$, the *discrete Fourier transform (DFT)* is given by

$$\widehat{f}(c)[j] := \sum_{k=0}^{n-1} c_k \exp\left(-\frac{2i\pi jk}{n}\right) \qquad \text{for } j = 0, \dots, n-1,$$

and the *inverse DFT* of c is given by

$$f(c)[k] := \frac{1}{n} \sum_{j=0}^{n-1} c_j \exp\left(\frac{2\mathrm{i}\pi jk}{n}\right) \qquad \text{for } k = 0, \ldots, n-1.$$

In general, both transforms require a computational effort of the order $\mathcal{O}(n^2)$, but the fast Fourier transform (FFT) algorithm by Cooley and Tukey [19] exploits the symmetry and periodicity of complex exponentials of the DFT and reduces the complexity of both transforms to $\mathcal{O}(n \log n)$.

Theorem B.3 *For* a, b $\in \mathbb{R}^n$, *the identity* $(a * b) = f(\widehat{f}(a) \bullet \widehat{f}(b))$ *holds, where* $\bullet$ *is the pointwise multiplication.*

This implies that the complexity of the discrete convolution reduces to $\mathcal{O}(n \log n)$ by FFT.

Algorithm B.4 *On the equidistant grid* $\mathcal{T} = \{t_i = \frac{iT}{n} : i = 0, 1, \ldots, n\}$:

1) *Draw i.i.d. random variables* $\zeta_{j,i} \sim \mathcal{N}(0, 1)$ *for* $j = 1, \ldots, M, i = 1, \ldots, n$.
2) *Define* $\mathfrak{g} := (g(t_i))_{i=1,\ldots,n}$ *and* $\zeta_j := (\zeta_{j,i})_{i=1,\ldots,n}$ *for* $j = 1, \ldots, M$.
3) *Using the FFT, compute* $\varphi_j := \widehat{f}(\mathfrak{g}) \bullet \widehat{f}(\zeta_j)$ *for* $j = 1, \ldots, M$.
4) *Simulate M paths of* $\mathcal{G}^\alpha W$, *using the FFT, as* $(\mathcal{G}^\alpha W)^j(\mathcal{T}) = \sqrt{\frac{T}{n}} f(\varphi_j)$ *for* $j = 1, \ldots, M$.

In Step 2 in Algorithm B.4, we may replace the evaluation points $\mathfrak{g}$ by any optimal evaluation point $(g(t_i^*))_{i=1}^n$ as in (4.2). Many packages offer a direct implementation of the discrete convolution such as `numpy.convolve` in `Python`. The user then only needs to pass the arguments $\mathfrak{g}$ and ξ_j, and Steps 3 and 4 in Algorithm B.4 are computed automatically (using efficient FFT techniques) by the function. Although the FFT step is the heaviest computation in the simulation of rough volatility models, the actual time grid $\mathcal{T}$ is not especially large ($n \ll 1000$). Hence the fastest FFT for very large n is not essential as the implementation is run on smaller time grids. In this aspect, we find that `numpy.convolve` is a very competitive implementation.

Appendix C: Proof of Proposition 2.2

In this section, we present a proof of Proposition 2.2. We first consider the case $\alpha \in (-\lambda, 0)$ with $0 < \lambda \leq 1$. Fix $f \in \mathcal{C}^\lambda(\mathbb{I})$ and $g \in \mathcal{L}^\alpha$. As our first step, we derive a useful representation akin to Samko et al. [65, Eq. (13.1)], but for the operator $\mathcal{G}^\alpha$ associated to g, which amounts to $\mathcal{G}^\alpha f(0) = 0$ and

$$\mathcal{G}^\alpha f(t) = \left(f(t) - f(0)\right)g(t) - \int_0^t \left(f(t) - f(s)\right)\frac{\mathrm{d}}{\mathrm{d}t}g(t-s)\mathrm{d}s \qquad (\text{C.1})$$

for $t \in (0, 1]$ (note that $g(0)$ need not be defined, but the assumptions on f and g yield $\mathcal{G}^\alpha f(0) = 0$ by its definition in (2.1)). To show that (C.1) holds, we look at

the difference quotients for the definition of $\mathcal{G}^\alpha f$ in (2.1). For any $t \in [0, 1)$ and any small enough $h > 0$, a bit of rewriting leads to the equality

$$\int_0^{t+h} \big(f(s) - f(0)\big)g(t + h - s)ds - \int_0^t \big(f(s) - f(0)\big)g(t - s)ds$$

$$= \int_0^t \big(f(s) - f(t)\big)\big(g(t + h - s) - g(t - s)\big)ds$$

$$+ \big(f(t) - f(0)\big)\left(\int_0^{t+h} g(t + h - s)ds - \int_0^t g(t - s)ds\right)$$

$$+ \int_t^{t+h} \big(f(s) - f(t)\big)g(t + h - s)ds. \tag{C.2}$$

In the second term on the right-hand side of (C.2), a change of variables gives

$$\big(f(t) - f(0)\big)\left(\int_0^{t+h} g(t+h-s)ds - \int_0^t g(t-s)ds\right) = \big(f(t) - f(0)\big)\int_{-h}^0 g(t-r)dr.$$

Looking at the third term on the right-hand side of (C.2), using the assumptions $f \in \mathcal{C}^\lambda$ and $g \in \mathcal{L}^\alpha$ for the given λ and α, we get

$$\left|\int_t^{t+h} \big(f(s) - f(t)\big)g(t + h - s)ds\right| \leq \int_t^{t+h} C_1 h^\lambda C_2 h^\alpha ds \leq C_0 h^{1+\lambda+\alpha} = o(h)$$

as h tends to zero, since $\lambda + \alpha \in (0, 1)$. As for the first term, we have

$$\left|\big(f(t) - f(s)\big)\frac{g(t + h - s) - g(t - s)}{h}\right| \leq \frac{C_1 |t - s|^\lambda}{h} \int_{t-s}^{t-s+h} g'(r)dr$$

$$\leq C_1(t - s)^\lambda C_2 |t - s|^{\alpha-1}$$

$$= C_0(t - s)^{\lambda+\alpha-1}$$

for all $s \in (0, t)$, $h > 0$, where $\lambda + \alpha - 1 \in (-1, 0)$; so the right-hand side is in $L^1([0, t])$ and we can apply the dominated convergence theorem. Specifically, dividing by h in (C.2) and sending h to zero, we obtain (C.1) in the limit by using dominated convergence on the first term, Lebesgue's differentiation theorem on the second, and noting that the third term vanishes.

Having established (C.1), we use it to obtain the desired Hölder estimates. We begin with the first term on the right-hand side of (C.1). Let $\phi(t) := (f(t) - f(0))g(t)$ for $t \in \mathbb{I}$, where we note that $|(f(t) - f(0))g(t)| \leq Ct^{\lambda+\alpha}$ with $\lambda + \alpha \in (0, 1)$ so that $\phi(0) = 0$ is well defined. Rewriting and using the assumptions $f \in \mathcal{C}^\lambda$ and $g \in \mathcal{L}^\alpha$ for the given λ and α, we get for every $t \in \mathbb{I}$ and $h \in (0, 1 - t]$ that

$$|\phi(t + h) - \phi(t)| \leq |f(t) - f(0)| \int_t^{t+h} |g'(r)|dr + |g(t + h)||f(t + h) - f(t)|$$

$$\leq C|f|_\lambda t^{\lambda+\alpha}(t + h)^\alpha\big((t + h)^{-\alpha} - t^{-\alpha}\big) + C|f|_\lambda h^{\lambda+\alpha}$$

$$\leq C'|f|_\lambda h^{\lambda+\alpha} \tag{C.3}$$

where the last inequality follows by elementary considerations as in the arguments in Muskhelishvili [56, Chap. 1, Sect. 6, item 5]. The case $h \in [-t, 0]$ is analogous.

For the second term on the right-hand side of (C.1), we can follow a procedure similar to the proof of Samko et al. [65, Lemma 13.1]. Defining

$$\varphi(t) := \int_0^t \big(f(t) - f(s)\big)\frac{\mathrm{d}}{\mathrm{d}t}g(t-s)\mathrm{d}s = \int_0^t \big(f(t) - f(t-r)\big)\frac{\mathrm{d}}{\mathrm{d}r}g(r)\mathrm{d}r$$

and rewriting things, we arrive for any $t \in \mathbb{I}$ and $h \in [-t, 1-t]$ at

$$\begin{aligned}
\varphi(t+h) - \varphi(t) &= \int_0^t \big(f(t) - f(t-r)\big)\big(g'(r+h) - g'(r)\big)\mathrm{d}r \\
&\quad - \int_0^t \big(f(t) - f(t-r)\big)g'(r+h)\mathrm{d}r \\
&\quad + \int_{-h}^t \big(f(t+h) - f(t-u)\big)g'(u+h)\mathrm{d}u \\
&= \int_0^t \big(f(t) - f(t-r)\big)\big(g'(r+h) - g'(r)\big)\mathrm{d}r \\
&\quad - \int_0^t \big(f(t+h) - f(t)\big)g'(r+h)\mathrm{d}r \\
&\quad + \int_{-h}^0 \big(f(t+h) - f(t-u)\big)g'(u+h)\mathrm{d}u \\
&=: I_1 + I_2 + I_3.
\end{aligned} \tag{C.4}$$

Without loss of generality, we assume $h > 0$. For the first integral, a change of variables gives

$$\begin{aligned}
|I_1| &\le |f|_\lambda \int_0^t r^\lambda \int_r^{r+h} |g''(u)|\mathrm{d}u\mathrm{d}r \\
&\le C|f|_\lambda \int_0^t r^\lambda \big(r^{\alpha-1} - (r+h)^{\alpha-1}\big)\mathrm{d}r \\
&= C|f|_\lambda h^{\lambda+\alpha-1} \int_0^t \Big(\frac{r}{h}\Big)^\lambda \Big(\Big(\frac{r}{h}\Big)^{\alpha-1} - \Big(\frac{r}{h}+1\Big)^{\alpha-1}\Big)\mathrm{d}r \\
&= C|f|_\lambda h^{\lambda+\alpha} \int_0^{t/h} u^{\alpha+\lambda-1}\Big(1 - \Big(1+\frac{1}{u}\Big)^{\alpha-1}\Big)\mathrm{d}u \\
&\le C|f|_\lambda h^{\lambda+\alpha}\Big(\int_0^1 u^{\lambda+\alpha-1}\mathrm{d}u + (1-\alpha)\int_1^\infty u^{\lambda+\alpha-2}\mathrm{d}u\Big),
\end{aligned}$$

where $\lambda + \alpha \in (0, 1)$; so the final two terms on the right-hand side are finite. In the final line, we have used that the mapping $y \mapsto -(1+y)^{\alpha-1} + (\alpha-1)y$ is concave with a maximum value of -1 at $y = 0$.

As regards the two remaining integrals I_2 and I_3, we see immediately that

$$|I_2| \leq C|f|_\lambda h^\lambda \int_0^\infty (r+h)^{\alpha-1} \mathrm{d}r = \frac{C}{|\alpha|}|f|_\lambda h^{\lambda+\alpha},$$

$$|I_3| \leq C|f|_\lambda \int_{-h}^0 (u+h)^{\lambda+\alpha-1} \mathrm{d}u = \frac{C}{|\lambda+\alpha|}|f|_\lambda h^{\lambda+\alpha}.$$

By linearity, the desired continuity of the operator $\mathcal{G}^\alpha : \mathcal{C}^\lambda(\mathbb{I}) \to \mathcal{C}^{\lambda+\alpha}(\mathbb{I})$ for $\alpha \in (-\lambda, 0)$ now follows from (C.1), (C.3) and the three above estimates for (C.4).

It remains to consider $\alpha \in (0, 1-\lambda)$. As before, recall $0 < \lambda \leq 1$ and fix $f \in \mathcal{C}^\lambda(\mathbb{I})$ along with $g \in \mathcal{L}^\alpha$. Unlike above, $s \mapsto \frac{\mathrm{d}}{\mathrm{d}t}g(t-s)$ is now integrable on $\mathbb{I}$ which makes things go through more easily; in particular, we can work directly with the definition of $\mathcal{G}^\alpha f$ in (2.1), applying arguments analogous to (C.4). The case $g(u) = u^\alpha$ is already covered by the proof of Theorem A.2. For a general $g \in \mathcal{G}^\alpha$, we can retrace those same steps except that in (A.2) and the subsequent estimates for J_1, J_2 and J_3, we must now invoke our control on g and its derivatives (similarly to how we did it above for (C.4) and the subsequent estimates of I_1, I_2 and I_3). This completes the proof of Proposition 2.2. $\qquad\square$

Declarations

References

1. Akyıldırım, E., Dolinsky, Y., Soner, H.M.: Approximating stochastic volatility by recombinant trees. Ann. Appl. Probab. **24**, 2176–2205 (2014)
2. Alòs, E., León, J., Vives, J.: On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. Finance Stoch. **11**, 571–589 (2007)
3. Bardina, X., Nourdin, I., Rovira, C., Tindel, S.: Weak approximation of a fractional SDE. Stoch. Process. Appl. **120**, 39–65 (2010)
4. Barndorff-Nielsen, O.E., Schmiegel, J.: Ambit processes: with applications to turbulence and tumour growth. In: Benth, F.E., et al. (eds.) Stochastic Analysis and Applications. Abel Symp., vol. 2, pp. 93–124. Springer, Berlin (2007)

5. Bayer, C., Friz, P., Gassiat, P., Martin, J., Stemper, B.: A regularity structure for rough volatility. Math. Finance **30**, 782–832 (2020)
6. Bayer, C., Friz, P., Gatheral, J.: Pricing under rough volatility. Quant. Finance **16**, 1–18 (2016)
7. Bayer, C., Friz, P., Gulisashvili, A., Horvath, B., Stemper, B.: Short-time near the money skew in rough fractional stochastic volatility models. Quant. Finance **19**, 779–798 (2019)
8. Beiglböck, M., Siorpaes, P.: Pathwise versions of the Burkholder–Davis–Gundy inequality. Bernoulli **21**, 360–373 (2015)
9. Bender, C., Parczewski, P.: On the connection between discrete and continuous Wick calculus with an application to the fractional Black–Scholes model. In: Cohen, S.N., et al. (eds.) Stochastic Processes, Finance and Control: A Festschrift in Honor of Robert J. Elliott, pp. 3–40. World Scientific, Singapore (2012)
10. Bennedsen, M., Lunde, A., Pakkanen, M.S.: Hybrid scheme for Brownian semistationary processes. Finance Stoch. **21**, 931–965 (2017)
11. Bennedsen, M., Lunde, A., Pakkanen, M.S.: Decoupling the short- and long-term behavior of stochastic volatility. J. Financ. Econom. **20**, 961–1006 (2022)
12. Bergomi, L.: Smile dynamics II. Risk, October, 67–73 (2005)
13. Billingsley, P.: Convergence of Probability Measures, 2nd edn. Wiley, New York (1999)
14. Boucheron, S., Lugosi, G., Massart, P.: Concentration Inequalities: A Nonasymptotic Theory of Independence. Oxford University Press, London (2013)
15. Broux, L., Caravenna, F., Zambotti, L.: Hairer's multilevel Schauder estimates without regularity structures. Preprint (2023). Available online at https://arxiv.org/abs/2301.07517
16. Čentsov, N.N.: Weak convergence of stochastic processes whose trajectories have no discontinuities of the second kind and the 'heuristic' approach to the Kolmogorov–Smirnov tests. Theory Probab. Appl. **1**, 140–144 (1956)
17. Clark, D.S.: Short proof of a discrete Gronwall inequality. Discrete Appl. Math. **16**, 279–281 (1987)
18. Comte, F., Renault, É.: Long memory continuous time models. J. Econom. **73**, 101–149 (1996)
19. Cooley, J.W., Tukey, J.W.: An algorithm for the machine calculation of complex Fourier series. Math. Comput. **19**, 297–301 (1965)
20. Djehiche, M., Eddahbi, M.: Hedging options in market models modulated by the fractional Brownian motion. Stoch. Anal. Appl. **19**, 753–770 (2001)
21. Donsker, M.D.: An invariance principle for certain probability limit theorems. In: Memoirs of the AMS, vol. 6. Am. Math. Soc., Providence (1951)
22. Drábek, P.: Continuity of Nemyckij's operator in Hölder spaces. Comment. Math. Univ. Carol. **16**, 37–57 (1975)
23. Dudley, R.M.: Uniform Central Limit Theorems. Cambridge University Press, Cambridge (1999)
24. El Euch, O., Rosenbaum, M.: The characteristic function of rough Heston models. Math. Finance **29**, 3–38 (2019)
25. Ethier, S.N., Kurtz, T.G.: Markov Processes: Characterization and Convergence. Wiley, New York (1986)
26. Forde, M., Zhang, H.: Asymptotics for rough stochastic volatility models. SIAM J. Financ. Math. **8**, 114–145 (2017)
27. Friz, P.K., Hairer, M.: A Course on Rough Paths. Springer, Berlin (2021)
28. Friz, P.K., Victoir, N.: Multidimensional Stochastic Processes as Rough Paths: Theory and Applications. Cambridge University Press, Cambridge (2010)
29. Fukasawa, M.: Asymptotic analysis for stochastic volatility: martingale expansion. Finance Stoch. **15**, 635–654 (2011)
30. Fukasawa, M., Takabatake, T., Westphal, R.: Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics (Is volatility rough?). Math. Finance **32**, 1086–1132 (2022)
31. Gassiat, P.: On the martingale property in the rough Bergomi model. Electron. Commun. Probab. **24**, 1–9 (2019)
32. Gatheral, J., Jaisson, T., Rosenbaum, M.: Volatility is rough. Quant. Finance **18**, 933–949 (2018)
33. Guennoun, H., Jacquier, A., Roome, P., Shi, F.: Asymptotic behaviour of the fractional Heston model. SIAM J. Financ. Math. **9**, 1017–1045 (2018)
34. Gulisashvili, A.: Large deviation principle for Volterra type fractional stochastic volatility models. SIAM J. Financ. Math. **9**, 1102–1136 (2018)
35. Hamadouche, D.: Invariance principles in Hölder spaces. Port. Math. **57**, 127–151 (2000)
36. Hardy, G.H., Littlewood, J.E.: Some properties of fractional integrals I. Math. Z. **27**, 565–606 (1928)
37. Heinrich, C., Pakkanen, M., Veraart, A.E.D.: Hybrid simulation scheme for volatility modulated moving average fields. Math. Comput. Simul. **166**, 224–244 (2019)

38. Horvath, B., Jacquier, A., Lacombe, C.: Asymptotic behaviour of randomised fractional volatility models. J. Appl. Probab. **56**, 496–523 (2019)
39. Hurst, H.E.: Long-term storage capacity of reservoirs. Trans. Am. Soc. Civ. Eng. **116**, 770–799 (1956)
40. Jacquier, A., Martini, C., Muguruza, A.: On VIX futures in the rough Bergomi model. Quant. Finance **18**, 45–61 (2018)
41. Jacquier, A., Pakkanen, M., Stone, H.: Pathwise large deviations for the rough Bergomi model. J. Appl. Probab. **55**, 1078–1092 (2018)
42. Jakubowski, A., Mémin, J., Pagès, G.: Convergence en loi des suites d'intégrales stochastiques sur l'espace D1 de Skorokhod. PTRF **81**, 111–137 (1989)
43. Karatzas, I., Shreve, S.E.: Methods of Mathematical Finance. Springer, New York (1998)
44. Kolmogorov, A.: Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. C.R. (Dokl.) Acad. URSS (N.S.) **26**, 115–118 (1940)
45. Krylov, N.V.: Lectures on Elliptic and Parabolic Equations in Hölder Spaces. Graduate Studies in Mathematics, vol. 12. Am. Math. Soc., Providence (1996)
46. Kurtz, T.G., Protter, P.E.: Weak convergence of stochastic integrals and stochastic differential equations. In: Talay, D., Tubaro, L. (eds.) CIME School in Probability. Lecture Notes in Mathematics, vol. 1627, pp. 1–41. Springer, Berlin (1991)
47. Kushner, J.: On the weak convergence of interpolated Markov chains to a diffusion. Ann. Probab. **2**, 40–50 (1974)
48. Lamperti, J.: Semi-stable stochastic processes. Trans. Am. Math. Soc. **104**, 62–78 (1962)
49. Lamperti, J.: On convergence of stochastic processes. Trans. Am. Math. Soc. **104**, 430–435 (1962)
50. Livieri, G., Mouti, S., Pallavicini, A., Rosenbaum, M.: Rough volatility: evidence from option prices. IISE Trans. **50**, 767–776 (2018)
51. Lyons, T.J.: Differential equations driven by rough signals. Rev. Mat. Iberoam. **14**, 215–310 (1998)
52. Mandelbrot, B., Van Ness, J.: Fractional Brownian motions, fractional noises and applications. SIAM Rev. **10**, 422–437 (1968)
53. Mann, H.B., Wald, A.: On stochastic limit and order relationships. Ann. Math. Stat. **14**, 217–226 (1943)
54. McCrickerd, R., Pakkanen, M.S.: Turbocharging Monte Carlo pricing for the rough Bergomi model. Quant. Finance **18**, 1877–1886 (2018)
55. Mulinacci, S., Pratelli, M.: Functional convergence of Snell envelopes: applications to American options approximations. Finance Stoch. **2**, 311–327 (1998)
56. Muskhelishvili, N.I.: Singular Integral Equations. Boundary Problems of Functions Theory and Their Applications to Mathematical Physics. Revised translation from the Russian, edited by J.R.M. Radok. Reprinted. Wolters-Noordhoff Publishing, Groningen (1972).
57. Neuenkirch, A., Shalaiko, T.: the order barrier for strong approximation of rough volatility models. Preprint (2016). Available online at https://arxiv.org/abs/1606.03854
58. Neuman, E., Rosenbaum, M.: Fractional Brownian motion with zero Hurst parameter: a rough volatility viewpoint. Electron. Commun. Probab. **23**, Paper no. 61, 1–12 (2018)
59. Nieminen, A.: Fractional Brownian motion and martingale-differences. Stat. Probab. Lett. **70**, 1–10 (2004)
60. Parczewski, P.: Donsker-type theorems for correlated geometric fractional Brownian motions and related processes. Electron. Commun. Probab. **22**, Paper no. 55, 1–13 (2017)
61. Parczewski, P.: A fractional Donsker theorem. Stoch. Anal. Appl. **32**, 328–347 (2014)
62. Pollard, D.: Convergence of Stochastic Processes. Springer, Berlin (1984)
63. Račkauskas, A., Suquet, C.: Necessary and sufficient condition for the functional central limit theorem in Hölder spaces. J. Theor. Probab. **17**, 221–243 (2004)
64. Romano, M., Touzi, N.: Contingent claims and market completeness in a stochastic volatility model. Math. Finance **7**, 399–412 (1997)
65. Samko, S.G., Kilbas, A.A., Marichev, O.I.: Fractional Integrals and Derivatives: Theory and Applications. Gordon & Breach, Yverdon (1993)
66. Sottinen, T.: Fractional Brownian motion, random walks and binary market models. Finance Stoch. **5**, 343–355 (2001)
67. Taqqu, M.: Weak convergence to fractional Brownian motion and to the Rosenblatt process. Z. Wahrscheinlichkeitstheor. Verw. Geb. **31**, 287–302 (1975)